DOCUMENT RESUME

ED 242 783 TM 840 201

AUTHOR Muraki, Eiji
TITLE Marginal Maximum Likelihood Estimation for Three-Parameter Polychotomous Item Response Models: Application of an EM Algorithm.
PUB DATE Apr 84
NOTE 70p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Algorithms; Data Analysis; *Estimation (Mathematics); Goodness of Fit; *Latent Trait Theory; *Mathematical Models; *Maximum Likelihood Statistics; Rating Scales; Research Design; Research Methodology
IDENTIFIERS *EM Algorithm; Graded Response Model; Item Parameters; Likert Scales; Rating Scale Model

ABSTRACT

This study examines the application of the marginal maximum likelihood (MML) EM algorithm to the parameter estimation problem of the three-parameter normal ogive and logistic polychotomous item response models. A three-parameter normal ogive model, the Graded Response model, has been developed on the basis of Samejima's two-parameter graded response model. A three-parameter logistic model, the Rating Scale model, has been developed based on Andrich's Rasch polychotomous response model. In the three-parameter models, an item response parameter is resolved into two parameters: item location and category (step). In the case of the Likert-type questionnaire, where only a single scale is employed to evoke different responses for test items, the three-parameter model is expected to be more useful in terms of prediction and analysis because the estimates of the slope and location parameters associated with the points on a single Likert scale can be separately estimated. The advantages of this type of model are demonstrated by analyzing actual data sets. An extension of measurement models of analytic models is also discussed. In a specific analytic model, based on the Graded Response Model, the structural item response model, the design matrix is incorporated to represent a structure among subject groups. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED242783

MARGINAL MAXIMUM LIKELIHOOD ESTIMATION FOR THREE-PARAMETER POLYCHOTOMOUS ITEM RESPONSE MODELS: APPLICATION OF AN EM ALGORITHM

Eiji Muraki
National Opinion Research Center
at the University of Chicago

A paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, 1984.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E. Muraki

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 840 201

Latent Trait Theory

Latent trait theory assumes that examinee performance on a test can be predicted or explained by defining examinee characteristics, referred to as traits, estimating scores for examinees on these traits, and using these scores to predict or explain test performance (Hambleton, et al., 1978, p.469). Since the traits are not directly observable, they are referred to as latent traits. An item response model, based on the latent trait theory, is used to express a probabilistic relationship between examinee performance on a test item and his latent trait.

The Purpose of the Study

The MML-EM algorithm has been actually applied to the estimation problem in the two-parameter normal ogive model (Bock & Aitkin, 1981) and the one-parameter logistic model (Thissen, 1982). The estimates of the parameters in Thurstone's (1947) multiple factor model are also obtained by this method (Bock & Aitkin, 1981). In this study, the

MML-EM algorithm is applied to the estimation problem of specific models, namely the three-parameter normal ogive model, which is a simple reformation of Samejima's Graded Response model, and a general case of Andrich's Rating Scale model. The main purpose of the study is to show that the Likert-type response data can be analyzed based on these models and that the MML-EM algorithm is also workable for the estimation problem.

The Graded Response Model

Supposed that any response item j is classified into m categories, 1 through m, the resulting response pattern $\underline{W}^*_i$ for a given subject is a sequence of integers. The response pattern for n graded items is given by the following form:

$$\underline{W}^*_i = (W^*_1, W^*_2, \ldots, W^*_n)_i \qquad (1)$$

where $1 \leq W^*_j \leq m$. From the assumption of ordered categories, the m possible categorical responses W^*_j to item j can be expressed by the following selection vectors $\underline{W}^*_j$:

Response to Item		1	2	3	.	.	.	m
$\underline{W}^*_j$	1	1	0	0	.	.	.	0
$\underline{W}^*_j$	2	1	1	0	.	.	.	0
.	.	.	.	.	.	.	.	.
$\underline{W}^*_j$	m	1	1	1	.	.	.	1

(2)

If we define $P^*_{j,k}(\theta)$ and $P^*_{j,k-1}(\theta)$ as the regression of the binary item score method in which all response categories less than k and k-1 are scored 1 for each item j

respectively, the operating characteristic (Samejima, 1972) of the graded item scoring for the latent trait variable θ is

$$P_{jk}(\theta) = P^*_{jk}(\theta) - P^*_{j,k-1}(\theta) \tag{3}$$

under the assumption that the discrimination parameters are equal for any $P^*_{jk}(\theta)$. Consequently, the matrix (2) can be rewritten as

Response to Item		1	2	3	.	.	.	m
$\underline{W}_j$	1	1	0	0	.	.	.	0
$\underline{W}_j$	2	0	1	0	.	.	.	0
. . .		.	.	.	.	.	.	.
$\underline{W}_j$	m	0	0	0	.	.	.	1

(4)

where $\underline{W}_j$ for k-th response is obtained by subtracting $\underline{W}^*_j$ for k-1 th response from $\underline{W}^*_j$ for k-th response.

Samejima (1969) further defines $P^*_{j0}(\theta)$ and $P^*_{jm}(\theta)$ so that

$$\begin{aligned} P^*_{j0}(\theta) &= 0 \\ P^*_{jm}(\theta) &= 1 \end{aligned} \tag{5}$$

Therefore,

$$\begin{aligned} P_{j1} &= P^*_{j1} - P^*_{j0} = P^*_{j1} \\ P_{j2} &= P^*_{j2} - P^*_{j1} \\ P_{jm} &= P^*_{jm} - P^*_{j,m-1} = 1 - P^*_{j,m-1} \end{aligned} \tag{6}$$

and

$$P_{jk} = P^*_{jk} - P^*_{j,k-1} \geq 0 \tag{7}$$

On the normal ogive model (Samejima, 1974), the formula for $P^*_{jk}(\theta)$ is given by

$$P^*_{jk}(\theta) = \int_{-\infty}^{a_j(\theta - b_{jk})} \phi(t)dt \tag{8}$$

where $\phi(t)=1/\sqrt{2\pi}\ \exp(-t^2/2)$, a_j is a common discrimination parameter, and b_{jk} is called a category threshold parameter. For each m-category item, there are m-1 category threshold parameters. The dichotomous level is, therefore, a special case of the graded level situation, in which m=2. Finally θ is a parameter which represents the latent trait of an individual subject.

The three-parameter normal ogive model is a direct extension of Samejima's two-parameter item response model, that is

$$P^*_{jk}(\theta) = \int_{-\infty}^{a_j(\theta-b_j+c_k)} \phi(t)dt \qquad (9)$$

Since my intention is to estimate these parameters for the Likert-type attitude data, I simply refer to a_j, b_j, c_k, and θ as the slope parameter, the location parameter, the category parameter, and the attitude parameter or latent trait variable respectively.

In the same way, we can write

$$P^*_{j,k-1}(\theta) = \int_{-\infty}^{a_j(\theta-b_j+c_{k-1})} \phi(t)dt \qquad (10)$$

Since $a_j>0$ then $P_{jk} = P^*_{jk} - P^*_{j,k-1} \geq 0$, then

$$c_k - c_{k-1} \geq 0 \qquad (11)$$

From (3), we can obtain the operating characteristic of the graded response category on the normal ogive model:

$$P_{jk}(\theta) = \int_{a_j(\theta-b_j+c_{k-1})}^{a_j(\theta-b_j+c_k)} \phi(t)dt \qquad (12)$$

The item response function, as mentioned earlier, can be viewed as the regression of item scores on the latent trait. The probability of a specific categorical response k to item j depends only on a given latent trait level θ_0, not on the number of people at θ_0, nor on the number of people at other latent trait levels θ_1, θ_2, . . . (Lord, 1980, p.34). Therefore, the parameters, a_j, b_j, and c_k are also invariant parameters.

Both (8) and (9) are the latent trait models for item responses which are scored in successive categories. A major distinction between these two models is that in the three-parameter normal ogive model (9) Samejima's parameter b_{jk} is resolved into an item location parameter b_j and a category parameter c_k.

If each questionnaire item has its own response categories, which may be different sizes m_j (j=1,2,...,n), Samejima's model (8) may be more appropriate than the three-parameter normal ogive model (9). For the Likert-type questionnaire, where only a single scale is employed to evoke different responses for test items, the three-parameter model is expected to be more useful in the prediction and analysis because the estimates of the parameters associated with items and the category parameters associated with the points on the single Likert scale can be separately estimated. The items can be unidimensionally ordered by a single measure, that is, their location estimates. The psy-

chological distance among points on the Likert scale can be estimated independently of the items.

The Rating Scale Model

Birnbaum's (Birnbaum, 1968) two-parameter logistic item response model is directly derived from Lord's normal ogive model on the basis of the approximation $|\Phi(x)-\Psi[1.7(x)]|<0.01$ for all x. Although the Rasch dichotomous model was developed independently of the latent trait models, the only difference between the Rasch dichotomous model (Rasch, 1960) and Birnbaum's two-parameter logistic model is the introduction of the assumption about the discriminating power of test items. These models share the following common form:

$$P_i(W_j=1|\theta) = \frac{\exp[a(\theta-b_j)]}{1+\exp[a(\theta-b_j)]} \qquad (13)$$

If the discriminating power a is assumed to vary among test items ($a=a_j$), the model (13) is called Birnbaum's two-parameter logistic model. If the discriminating power a is assumed to be common for all items, the parameter a can be dropped from the model by arbitrarily determining a=1, and the model is called Rasch's dichotomous response model.

The Partial Credit model is formulated based on the assumption that each probability of choosing the k-th category over the k-1 th category is governed by the dichotomus response model. To develop the Partial Credit model, let us denote P_{jk} as the specific probability of choosing the k-th

category from m_j possible categories of item j. In the dichotomous model (m_j=2),

$$P_{j1} + P_{j2} = 1$$

The conditional probability of choosing category 2 given the probability of choosing categories 1 and 2 is, then,

$$P_{j2|12} = \frac{P_{j2}}{P_{j1}+P_{j2}} = P_{j2} = \frac{\exp[a_j(\theta-b_{j2})]}{1+\exp[a_j(\theta-b_{j2})]} \qquad (14)$$

Therefore,

$$P_{j1|12} = 1 - P_{j2} = \frac{1}{1+\exp[a_j(\theta-b_{j2})]}$$

In the polychotomous response model, in which m_j is more than 2 for item j,

$$P_{j1} + P_{j2} + \ldots + P_{jm} = \sum_{k=1}^{m_j} P_{jk} = 1 \qquad (15)$$

For each of adjacent categories, the probability of the specific categorical response k over k-1 is given by the conditional probability, which is the same as (14):

$$C_{jk} = P_{jk|k-1,k} = \frac{P_{jk}}{P_{j,k-1}+P_{jk}} = \frac{\exp[a_j(\theta-b_{jk})]}{1+\exp[a_j(\theta-b_{jk})]} \qquad (16)$$

Equation (16) becomes, then,

$$P_{jk} = \frac{C_{jk}}{1-C_{jk}} P_{j,k-1} \qquad (17)$$

where

$$\frac{C_{jk}}{1-C_{jk}} = \frac{P_{jk|k-1,k}}{1-P_{jk|k-1,k}} = \frac{P_{jk|k-1,k}}{P_{j,k-1|k-1,k}}$$

$$= \exp[a_j(\theta-b_{jk})] \qquad (18)$$

Equation (17) may be called the operating characteristic for the Partial Credit model. If we start by determining

$$P_{j1} = 1/G$$

we obtain the following probabilities by applying the operating characteristic (17):

$$P_{j2} = \exp[a_j(\theta-b_{j2})]/G$$

$$P_{j3} = \exp[a_j(\theta-b_{j2})+a_j(\theta-b_{j3})]/G$$

. . . . . .

$$P_{jg} = \exp[\textstyle\sum_{v=2}^{g} a_j(\theta-b_{jv})]/G$$

. . . . . .

$$P_{jm_j} = \exp[\textstyle\sum_{v=2}^{m} a_j(\theta-b_{jv})]/G$$

where g is a subscription for a specific categorical response k=g. Since $\sum_{k=1}^{m_j} P_{jk}=1$ shown in (15),

$$G = 1 + \sum_{c=2}^{m_j} \exp[\sum_{v=2}^{c} a_j(\theta-b_{jv})]$$

Therefore, the Partial Credit model is formulated by

$$P_{jk}(\theta) = \frac{\exp[\sum_{v=2}^{k} a_j(\theta-b_{jv})]}{1+\sum_{c=2}^{m_j} \exp[\sum_{v=2}^{c} a_j(\theta-b_{jv})]}$$

$$= \frac{\exp[\sum_{v=1}^{k} a_j(\theta-b_{jv})]}{\sum_{c=1}^{m_j} \exp[\sum_{v=1}^{c} a_j(\theta-b_{jv})]} \qquad (19)$$

where $\theta-b_{j1}=0$.

Masters (1980) calls the parameter b_{jk} in (19) an item step parameter. The b_{jk} is an inflexion point expressed by the conditional probability C_{jk} in (16). At this point ($\theta=b_{jk}$), $C_{jk}=0.5$. Thus, the parameter b_{jk} indicates the location of the logistic ogive curve of C_{jk} on the θ-scale. On the other hand,

if $\theta = b_{jk}$ $P_{jk} = P_{j,k-1}$

$\theta > b_{jk}$ $P_{jk} > P_{j,k-1}$

and $\theta < b_{jk}$ $P_{jk} < P_{j,k-1}$

under the assumption $a_j > 0$. Therefore, b_{jk} is a threshold parameter. It should be noted that b_{jk} is not sequentially ordered within item j because the parameter represents the relative magnitude of the adjacent probabilities P_{jk} and $P_{j,k-1}$. Furthermore, when all probabilities P_{jk} are equal, the values of the parameter also become identical.

The parameter a_j is a slope parameter which is common within each item. If the parameter is assumed to be equal for all items, the model (19) becomes Masters' Partial Credit model. The range of a_j is generally assumed to be from 0 to ∞ ($0<a_j<\infty$) from the same reason as was discussed in Samejima's model.

The Rating Scale model is the simple extension of the Partial Credit model, that is

$$P_{jk}(\theta) = \frac{\exp[\sum_{v=1}^{k} a_j(\theta-b_j+d_v)]}{\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} a_j(\theta-b_j+d_v)]} \qquad (20)$$

where $\theta-b_j+d_1=0$ and the parameter b_{jk} is resolved into two parameters b_j and d_k ($b_{jk}=b_j-d_k$).

Masters (1980) calls the parameters b_j and d_k in (20) a location parameter and a threshold parameter respectively. Since the values of the parameters (b_{jk}) are not necessarily ordered within item j, the threshold parameters (d_v) are not sequentially ordered for k=1,2,...,m. The parameter d_k is interpreted as the relative difficulty of step k in comparing other steps. For this reason, it is quite reasonable to simply call the parameter d_k a step parameter, distinguish-

ing the category parameter c_k in Samejima's model. For the same reason stated in the Partial Credit model, when all probabilities P_{jk} are equal, the values of the parameters (d_k) become the same. The parameter a_j and b_j are called a slope parameter and a location parameter respectively. These parameters are essentially interpretable in the same manner as the parameters in Samejima's model.

Item Parameter Estimation

Let us define an entry in the matrix (4) as (W_{jk}) (W_{jk}=1 or 0). Equation (12) is the probability function of W_{jk}=1 for the k-th category of item j, conditional on the latent trait variable θ_i. By the principle of local independence (Birnbaum, 1968), the conditional probability of a certain response pattern i, given θ, for m categories and n items, which is denoted by a response matrix, $W_i=(W_{jk})_i$, is the joint probability:

$$P_i((W_{jk})|\theta) = \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{jk}(\theta)]^{W_{jki}} \qquad (21)$$

For subjects randomly sampled from a population with a normal distribution of the latent trait variable, $\phi(\theta)$, the unconditional probability of the observed pattern i is the marginal distribution function:

$$P_i((W_{jk})) = \int_{-\infty}^{\infty} P_i((W_{jk})|\theta)\phi(\theta)d\theta \qquad (22)$$

If a subject i responds to n items with m categories, his response pattern i can be then assigned to one of m^n mutually exclusive patterns. Let the number of subjects observed in such a pattern i be r_i and its probability p_i. The r_i's are positive integer and each $p_i \geq 0$ for $i=1,2,\ldots,m^n$, and

$$\sum_{i=1}^{m^n} p_i = 1 \tag{23}$$

$$r_i = Np_i \tag{24}$$

$$E(p_i) = P_i \tag{25}$$

where N is the total number of subjects sampled from the population. The r_i is, therefore, multinomially distributed with parameters N and P_i. This probability can be interpreted as the likelihood function of the parameter, a_j, b_j, and c_k or d_k. The likelihood function L is

$$L = (N! \,/ \prod_{i=1}^{m^n} r_i!) \prod_{i=1}^{m^n} [P_i((W_{jk}))]^{r_i} \tag{26}$$

Taking the natural logarithm of (26), we obtain

$$\ln L = \ln N! - \ln \sum_{i=1}^{m^n} r_i! + \sum_{i=1}^{m^n} r_i \ln P_i((W_{jk})) \tag{27}$$

The maximum likelihood estimates $\hat{a}_j$, $\hat{b}_j$, and $\hat{c}_k$ or $\hat{d}_k$ can be derived from the first derivative of (27) with respect to each parameter, and respectively setting them to zero.

With respect to a_h, which is the parameter a_j for the specific item j=h, the likelihood equation (27) can be then differentiated as

$$\frac{\partial \ln L}{\partial a_h}$$

$$= \frac{\partial \sum_i^{m^n} r_i \ln P_i((W_{jk}))}{\partial a_h}$$

$$= \sum_i^{m^n} \frac{r_i}{P_i} \frac{\partial P_i((W_{jk}))}{\partial a_h}$$

$$= \sum_i^{m^n} \frac{r_i}{P_i} \frac{\partial [\int_{-\infty}^{\infty} \Pi_j^n \Pi_k^m [P_{jk}(\theta)]^{W_{jki}} \phi(\theta) d\theta]}{\partial a_h} \qquad (28)$$

since the first two terms of (27) are invariant with respect to a_h. Since the function $\phi(\theta)$ does not contain the parameter a_h, we only have to consider the function $[P_{jk}(\theta)]^{W_{jki}}$. Therefore,

$$\frac{\partial \Pi_j^n \Pi_k^m [P_{jk}(\theta)]^{W_{jki}}}{\partial a_h}$$

$$\frac{\Pi_j^n \Pi_k^m [P_{jk}(\theta)]^{W_{jki}}}{\Pi_k^m [P_{hk}(\theta)]^{W_{hki}}} \frac{\partial \Pi_k^m [P_{hk}(\theta)]^{W_{hki}}}{\partial a_h}$$

$$= \prod_j^n \prod_k^m [P_{jk}(\theta)]^{W_{jki}} \sum_k^m \frac{\partial [P_{hk}(\theta)]^{W_{hki}}}{\partial a_h} \frac{1}{[P_{hk}(\theta)]^{W_{hki}}} \qquad (29)$$

because

$$\frac{\partial \Pi_k^m [P_{hk}(\theta)]^{W_{hki}}}{\partial a_h}$$

$$= \frac{\partial [P_{h1}(\theta)]^{W_{h1i}}}{\partial a_h} \prod_{k=2}^m [P_{hk}(\theta)]^{W_{hki}} + \frac{\partial [P_{h2}(\theta)]^{W_{h2i}}}{\partial a_h} \cdot$$

$$\prod_{k \neq 2}^m [P_{hk}(\theta)]^{W_{hki}} + \cdots + \frac{\partial [P_{hm}(\theta)]^{W_{hmi}}}{\partial a_h} \prod_{k=1}^{m-1} [P_{hk}(\theta)]^{W_{hki}}$$

$$= \prod_{k=1}^m [P_{hk}(\theta)]^{W_{hki}} \left[\frac{\partial [P_{h1}(\theta)]^{W_{h1i}}}{\partial a_h} \frac{1}{[P_{h1}(\theta)]^{W_{h1i}}} \right.$$

$$+ \frac{\partial [P_{h2}(\theta)]^{W_{h2i}}}{\partial a_h} \frac{1}{[P_{h2}(\theta)]^{W_{h2i}}} + \cdots + \frac{\partial [P_{hm}(\theta)]^{W_{hmi}}}{\partial a_h} \cdot$$

$$\left. \frac{1}{[P_{hm}(\theta)]^{W_{hmi}}} \right]$$

From (28) and (29), we finally obtain

$$\frac{\partial \ln L}{\partial a_h}$$

$$= \sum_{i}^{s} \frac{r_i}{P_i} \int_{-\infty}^{\infty} \prod_{jk}^{nm} [P_{jk}(\Theta)]^{W_{jki}} \sum_{k} \frac{\partial [P_{hk}(\theta)]^{W_{hki}}}{\partial a_h} \frac{\phi(\theta) d\theta}{[P_{hk}(\theta)]^{W_{hki}}} \quad (30)$$

Using the Gauss-Hermite Quadrature, we can approximate (30) such that

$$\frac{\partial \ln L}{\partial a_h}$$

$$\approx \sum_{i}^{n} \frac{r_i}{P_i} \sum_{f}^{F} \left[\prod_{jk}^{nm} [P_{jk}(X_f)]^{W_{jki}} \sum_{k} \frac{\partial [P_{hk}(X_f)]^{W_{hki}}}{\partial a_h} \frac{A(X_f)}{[P_{hk}(X_f)]^{W_{hki}}} \right] \quad (31)$$

where $A(X_f)$ is the weight of the Gauss-Hermite Quadrature and X_f is the quadrature point (Stroud & Secrest, 1966). The quadrature weight, $A(X_f)$, in (31) is approximately the normal probability density of a N(0,1) random variable at the point X_f, such that $\sum_{f}^{F} A(X_f)=1$.

Now let the observed score patterns be indexed by $l=1,2,\ldots,s$ where $s \leq \min(N, m^n)$. The number of subjects who responded in pattern l, then, may be denoted by r_l, whose

$$\sum_{l}^{s} r_l = N \quad (32)$$

Any quantities subscripted by i are, therefore, transformed to the l-th subscripts. The maximum likelihood equation (31) becomes

$$\sum_{l}^{s} \frac{r_l}{P_l} \sum_{f}^{F} \left[\prod_{jk}^{nm} [P_{jk}(X_f)]^{W_{jkl}} \sum_{k} \frac{\partial [P_{hk}(X_f)]^{W_{hkl}}}{\partial a_h} \frac{A(X_f)}{[P_{hk}(X_f)]^{W_{hkl}}} \right]$$

$$= \sum_{l}^{s} \sum_{f}^{F} \frac{r_l L_l A(X_f)}{\tilde{P}_l} \sum_{k} \frac{\partial [P_{hk}(X_f)]^{W_{hkl}}}{\partial a_h} \frac{1}{[P_{hk}(X_f)]^{W_{hkl}}} \quad (33)$$

where

$$\tilde{P}_l = \sum_{fjk}^{Fnm} \prod [P_{jk}(X_f)]^{W_{jkl}} A(X_f) \quad (34)$$

and

$$L_l = \prod_{jk}^{nm} [P_{jk}(X_f)]^{W_{jkl}} \qquad (35)$$

The W_{hkl} can take only two possible values, 1 or 0. When $W_{hkl}=0$, $\partial[P_{hk}(X_f)]^{W_{hkl}}/\partial a_h=0$. On the other hand, when $W_{hkl}=1$, $[\partial[P_{hk}(X_f)]^{W_{hkl}}/\partial a_h]\,[P_{hk}(X_f)]^{-W_{hkl}} = [\partial P_{hk}(X_f)/\partial a_h]\,[P_{hk}(X_f)]^{-1}$. Therefore, Equation (33) can be rewritten as

$$\sum_{fkl}^{fms} \frac{r_l L_l A(X_f) W_{hkl}}{P_l} \frac{\partial P_{hk}(X_f)}{\partial a_h} \frac{1}{P_{hk}(X_f)}$$

$$= \sum_{fk}^{fm} \frac{\bar{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial a_h} \qquad (36)$$

where

$$\bar{r}_{hkf} = \frac{r_l L_l A(X_f) W_{hkl}}{P_l} \qquad (37)$$

Following the same procedure, we can straightforwardly obtain the first derivative of the maximum likelihood equation with respect to b_h, which is the parameter b_j for the specific item j=h. That is,

$$\frac{\partial \ln L}{\partial b_h} = \sum_{fk}^{fm} \frac{\bar{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial b_h} \qquad (38)$$

Bock and Aitkin (1981) applied the EM algorithm (Dempster, et al., 1977) to estimate the parameters for each item individually and repeated the iteration process over the n items until the estimates for all items became stable to the required number of places. According to Thissen (1982, p.181), this maximization technique may be called "univariate maximization" in contrast to "multivariate maximization", which requires the generation and inversion of the 2n x 2n information matrix. In the initial cycles of the

univariate maximization, the rate of convergence is slower than in the multivariate maximization, but the increased speed of each cycle makes up for that (Thissen, 1982, p.181). By applying the same iterative process to the polychotomous model, we obtain the following derivatives:

$$t_{a_h} = \sum_{\ell}^{r} \sum_{k}^{m} \frac{\bar{r}_{hk\ell}}{P_{hk}(X_\ell)} \frac{\partial P_{hk}(X_\ell)}{\partial a_h}$$

$$t_{a_h} = \sum_{\ell}^{r} \sum_{k}^{m} \frac{\bar{r}_{hk\ell}}{P_{hk}(X_\ell)} \frac{\partial P_{hk}(X_\ell)}{\partial b_h}$$

$$\frac{\partial^2 \ln L}{\partial a_h^2}$$

$$\approx \frac{\partial}{\partial a_h}\left[\sum_{\ell}^{r} \sum_{k}^{m} \frac{\bar{r}_{hk\ell}}{P_{hk}(X_\ell)} \frac{\partial P_{hk}(X_\ell)}{\partial a_h}\right]$$

$$= \sum_{\ell}^{r} \bar{N}_\ell \sum_{k}^{m}\left[-\frac{p_{hk\ell}}{P_{hk}^2(X_\ell)}\left[\frac{\partial P_{hk}(X_\ell)}{\partial a_h}\right]^2 + \frac{p_{hk\ell}}{P_{hk}(X_\ell)} \frac{\partial^2 P_{hk}(X_\ell)}{\partial a_h^2}\right] \quad (39)$$

$$\frac{\partial^2 \ln L}{\partial b_h^2}$$

$$\approx \sum_{\ell}^{r} \bar{N}_\ell \sum_{k}^{m}\left[-\frac{p_{hk\ell}}{P_{hk}^2(X_\ell)}\left[\frac{\partial P_{hk}(X_\ell)}{\partial b_h}\right]^2 + \frac{p_{hk\ell}}{P_{hk}(X_\ell)} \frac{\partial^2 P_{hk}(X_\ell)}{\partial b_h^2}\right] \quad (40)$$

$$\frac{\partial^2 \ln L}{\partial a_h \partial b_h}$$

$$\approx \sum_{\ell}^{r} \bar{N}_\ell \sum_{k}^{m}\left[-\frac{p_{hk\ell}}{P_{hk}^2(X_\ell)} \frac{\partial P_{hk}(X_\ell)}{\partial a_h} \frac{\partial P_{hk}(X_\ell)}{\partial b_h} + \frac{p_{hk\ell}}{P_{hk}(X_\ell)} \frac{\partial^2 P_{hk}(X_\ell)}{\partial a_h \partial b_h}\right] \quad (41)$$

After replacing the observed $p_{hk\ell}$ with its expected value, $P_{hk}(X_\ell)$, we obtain

$$V_{a_h a_h}$$

$$= E\left(-\frac{\partial^2 \ln L}{\partial a_h^2}\right)$$

$$= \sum_{\ell}^{r} \bar{N}_\ell \sum_{k}^{m} \frac{1}{P_{hk}(X_\ell)} \left[\frac{\partial P_{hk}(X_\ell)}{\partial a_h}\right]^2 \quad (42)$$

$$Vb_hb_h$$

$$= E\left(-\frac{\partial^2 \ln L}{\partial b_h^2}\right)$$

$$= \sum_f^F \bar{N}_f \sum_k^m \frac{1}{P_{hk}(X_f)} \left[\frac{\partial P_{hk}(X_f)}{\partial b_h}\right]^2 \qquad (43)$$

and finally

$$Va_hb_h$$

$$= E\left(-\frac{\partial^2 \ln L}{\partial a_h \partial b_h}\right)$$

$$= \sum_f^F \bar{N}_f \sum_k^m \frac{1}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial a_h} \frac{\partial P_{hk}(X_f)}{\partial b_h} \qquad (44)$$

since the rightmost term in (39), (40), and (41) vanishes under the summation on k. $\bar{N}_f$ is called the provisional expected sample size at the quadrature point f. $\bar{N}_f$ is computed by

$$\bar{N}_f = \sum_1^s [r_l L_l A(X_f)/\tilde{P}_l] \qquad (45)$$

Applying the Newton-Raphson method to any q-th stage estimates of the parameters reduces to

$$\begin{bmatrix}\hat{a}_h \\ \hat{b}_h\end{bmatrix}_{q+1} = \begin{bmatrix}\hat{a}_h \\ \hat{b}_h\end{bmatrix}_q + \begin{bmatrix}Va_ha_h & Va_hb_h \\ Va_hb_h & Vb_hb_h\end{bmatrix}^{-1} \begin{bmatrix}t_{a_h} \\ t_{b_h}\end{bmatrix} \qquad (46)$$

The rigorous proof about the approximations of the second derivatives (42), (43), and (44) by the product of the first derivatives is given by Kendall and Stuart (1973, pp.8-9). The matrix V in (46) is called the information. The vector $\underline{t}$ in (46) is called the gradient. Because the information matrix is always symmetric in this approximation, it is relatively easy to invert the matrix. The theo-

retical derivation of the Newton-Raphson method is also briefly mentioned by Kendall and Stuart (1973, p.57).

In the Graded Response model, $P_{jk}(X_f)$ is a cumulative standard normal distribution function, defined by $P_{jk}(X_f) = P^*_{jk}(X_f) - P^*_{j,k-1}(X_f)$. The first derivative of $P^*_{jk}(X_f)$ with respect to a or b is derived in the following manner: if $P^*_{jk}(X_f)$ is denoted by $\Phi[Z_{jk}(X_f)]$ and the parameter a or b is represented by ξ,

$$\frac{\partial \Phi[Z_{jk}(X_f)]}{\partial \xi} = \phi[Z_{jk}(X_f)] \frac{\partial Z_{jk}(X_f)}{\partial \xi} \tag{47}$$

where $\phi[Z_{jk}(X_f)]$ is the standard normal density and $Z_{jk}(X_f) = a_j(X_f - b_j + c_k)$ is the normal deviate.

In the Rating Scale model, $P_{jk}(X_f)$ is a logistic function so that the evaluation of several functions stated here becomes relatively simple in comparison with the normal ogive model. Equation (34) and (35) are written by

$$\begin{aligned}
\widetilde{P}_l &= \sum_{f=1}^{F} \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{jk}(X_f)]^{W_{jkl}} A(X_f) \\
&= \sum_{f=1}^{F} \prod_{j=1}^{n} \prod_{k=1}^{m} \left[\frac{\exp[\sum_{v=1}^{k} Z_{jv}(X_f)]}{\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_f)]}\right]^{W_{jkl}} A(X_f) \\
&= \sum_{f=1}^{F} \frac{\exp[\sum_{j=1}^{n} \sum_{k=1}^{m} W_{jkl} \sum_{v=1}^{k} Z_{jv}(X_f)]}{\prod_{j=1}^{n} \sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_f)]} A(X_f)
\end{aligned} \tag{48}$$

$$\begin{aligned}
L_l &= \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{jk}(X_f)]^{W_{jkl}} \\
&= \prod_{j=1}^{n} \prod_{k=1}^{m} \left[\frac{\exp[\sum_{v=1}^{k} Z_{jv}(X_f)]}{\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_f)]}\right]^{W_{jkl}} \\
&= \frac{\exp[\sum_{j=1}^{n} \sum_{k=1}^{m} W_{jkl} \sum_{v=1}^{k} Z_{jv}(X_f)]}{\prod_{j=1}^{n} \prod_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_f)]}
\end{aligned} \tag{49}$$

where $Z_{jv}(X_f) = a_j(X_f - b_j + d_v)$. Equation (48) and (49) are, then, used to evaluate $\bar{r}_{jkf}$ and $\bar{N}_f$ shown in (37) and (45).

The gradient and the information matrix in the Newton-Raphson method are obtained from (36), (38), (42), (43), and (44). The first derivatives with respect to parameters a_h and b_h are given as follows:

$$\frac{\partial P_{hk}(X_f)}{\partial a_h}$$

$$= \frac{\partial}{\partial a_h}\left[\frac{\exp[\sum_{v=1}^{k} Z_{hv}(X_f)]}{\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{hv}(X_f)]}\right]$$

$$= \frac{1}{[\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{hv}(X_f)]]^2}\left[\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{hv}(X_f)]\frac{\partial[\exp[\sum_{v=1}^{k} Z_{hv}(X_f)]]}{\partial a_h}\right.$$

$$\left. - \exp[\sum_{v=1}^{k} Z_{hv}(X_f)]\frac{\partial[\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{hv}(X_f)]]}{\partial a_h}\right]$$

$$= P_{hk}(X_f)\left[\sum_{v=1}^{k}(X_f - b_h + d_v) - \sum_{c=1}^{m} P_{hc}(X_f)\sum_{v=1}^{c}(X_f - b_h + d_v)\right]$$

because

$$\frac{\partial[\exp[\sum_{v=1}^{k} Z_{hv}(X_f)]]}{\partial a_h}$$

$$= \exp[\sum_{v=1}^{k} a_h(X_f - b_h + d_v)]\sum_{v=1}^{k}(X_f - b_h + d_v)$$

and

$$\frac{\partial[\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{hv}(X_f)]}{\partial a_h}$$

$$= \sum_{c=1}^{m}[\exp[\sum_{v=1}^{c} a_h(X_f - b_h + d_v)]\sum_{v=1}^{c}(X_f - b_h + d_v)]$$

In the same manner, we obtain

$$\frac{\partial P_{hk}(X_f)}{\partial b_h}$$

$$= P_{hk}(X_f)a_h[-k+\sum_{c=1}^{m} cP_{hc}(X_f)]$$

The gradient vector and the information matrix in the Newton-Raphson method are, then, evaluated by the following components:

$$t_{ah} = \sum_{f=1}^{F}\sum_{k=1}^{m}\bar{r}_{hkf}[\sum_{v=1}^{k}(X_f-b_h+d_v)-\sum_{c=1}^{m}P_{hc}(X_f)\sum_{v=1}^{c}(X_f-b_h+d_v)] \quad (50)$$

$$t_{bh} = \sum_{f=1}^{F}\sum_{k=1}^{m}\bar{r}_{hkf}a_h[-k+\sum_{c=1}^{m}cP_{hc}(X_f)] \quad (51)$$

$$Va_ha_h = \sum_{f=1}^{F}\bar{N}_f\sum_{k=1}^{m}P_{hk}(X_f)[\sum_{v=1}^{k}(X_f-b_h+d_v)-\sum_{c=1}^{m}P_{hc}(X_f)\sum_{v=1}^{c}(X_f-b_h+d_v)]^2 \quad (52)$$

$$Vb_hb_h = \sum_{f=1}^{F}\bar{N}_f\sum_{k=1}^{m}P_{hk}(X_f)a_h^2[-k+\sum_{c=1}^{m}cP_{hc}(X_f)]^2 \quad (53)$$

and

$$Va_hb_h = \sum_{f=1}^{F}\bar{N}_f\sum_{k=1}^{m}P_{hk}(X_f)a_h[\sum_{v=1}^{k}(X_f-b_h+d_v)-\sum_{c=1}^{m}P_{hc}(X_f)\sum_{v=1}^{c}(X_f-b_h+d_v)]\cdot [-k+\sum_{c=1}^{m}cP_{hc}(X_f)] \quad (54)$$

where $P_{hk}(X_f)$ and $P_{hc}(X_f)$ are evaluated by (20), and $d_1=0$.

Category Parameter Estimation

Since the category parameter c_g, which is the parameter c_k for the specific category $k=g$, is contained in both $P_{j,g}(\theta)$ and $P_{j,,g+1}(\theta)$ as shown in Equation (7), the first derivative of maximum likelihood function with respect to c_g can be derived by the following process:

$$\frac{\partial\prod_j\prod_k[P_{jk}(\theta)]^{N_jk}}{\partial c_g}$$

$$= \frac{\prod_j^n \prod_k^m [P_{jk}(\theta)]^{W_{jki}}}{\prod_j^n [P_{jg}(\theta)]^{W_{jgi}}} \frac{\partial \prod_j^n [P_{jg}(\theta)]^{W_{jgi}}}{\partial c_g} + \frac{\prod_j^n \prod_k^m [P_{jk}(\theta)]^{W_{jki}}}{\prod_j^n [P_{j,g+1}(\theta)]^{W_{j,g+1,i}}} \cdot$$

$$\frac{\partial \prod_j^n [P_{j,g+1}(\theta)]^{W_{j,g+1,i}}}{\partial c_g}$$

$$= \prod_j^n \prod_k^m [P_{jk}(\theta)]^{W_{jki}} \sum_j^n \left[\frac{\partial [P_{jg}(\theta)]^{W_{jgi}}}{\partial c_g} \frac{1}{[P_{jg}(\theta)]^{W_{jgi}}} \right.$$

$$\left. + \frac{\partial [P_{j,g+1}(\theta)]^{W_{j,g+1,i}}}{\partial c_g} \frac{1}{[P_{j,g+1}(\theta)]^{W_{j,g+1,i}}} \right]$$

$$= \prod_j^n \prod_k^m [P_{jk}(\theta)]^{W_{jki}} \sum_j^n \left[\frac{W_{jgi}}{P_{jg}(\theta)} - \frac{W_{j,g+1,i}}{P_{j,g+1}(\theta)} \right] \frac{\partial P_{jg}^*(\theta)}{\partial c_g} \quad (55)$$

because $W_{jgi}=1$ or 0, $W_{j,g+1}=1$ or 0, $\partial P_{jg}(\theta)/\partial c_g = \partial P_{jg}^*(\theta)/\partial c_g$, and $\partial P_{j,g+1}(\theta)/\partial c_g = -\partial P_{j,g+1}^*(\theta)/\partial c_g$. Then, the maximum likelihood equation with respect to c_g is

$$\frac{\partial \ln L}{\partial c_g}$$

$$= \sum_i^n \frac{r_i}{P_i} \int_{-\infty}^{\infty} \prod_j^n \prod_k^m [P_{jk}(\theta)]^{W_{jki}} \sum_j^n \left[\frac{W_{jgi}}{P_{jg}(\theta)} - \frac{W_{j,g+1,i}}{P_{j,g+1}(\theta)} \right] \frac{\partial P_{jg}^*(\theta)}{\partial c_g} \cdot$$

$$\phi(\theta) d\theta$$

$$\approx \sum_i^n \frac{r_i}{P_i} \sum_t^r \left[\prod_j^n \prod_k^m P_{jk}(X_t) \right]^{W_{jki}} \sum_j^n \left[\frac{W_{jgi}}{P_{jg}(X_t)} - \frac{W_{j,g+1,i}}{P_{j,g+1}(X_t)} \right] \frac{\partial P_{jg}^*(X_t)}{\partial c_g} \cdot$$

$$A(X_t)] \quad (56)$$

Equation (56) can be finally rewritten as

$$^t c_g \doteq \sum_t^r \sum_j^n \left[\frac{\bar{r}_{jgt}}{P_{jg}(X_t)} - \frac{\bar{r}_{j,g+1,t}}{P_{j,g+1}(X_t)} \right] \frac{\partial P_{jg}^*(X_t)}{\partial c_g} \quad (57)$$

It is not common for the Likert scale of more than eight categories to be used in psychological or sociological research. If the number of categories is relatively small, say $m \leq 12$, the Newton-Raphson method can be applied to estimate the parameters without difficulty. Furthermore,

since the covariance between categories of more than a two-point distance is simply zero, the information matrix becomes a (m-1)x(m-1) tridiagonal symmetric matrix. The generation and inversion of this matrix is relatively simple. The expectancy of the second derivatives of the maximum likelihood equation with respect to the category parameters, say c_g and $c_{g'}$, are

$$V_{c_g c_g} = E\left(-\frac{\partial^2 \ln L}{\partial c_g^2}\right) = \sum_r \bar{N}_r \sum_j \left[\frac{1}{P_{j,g}(X_r)} + \frac{1}{P_{j,g+1}(X_r)}\right]\left[\frac{\partial P_{j,g}^*(X_r)}{\partial c_g}\right]^2 \quad (58)$$

$$V_{c_g, c_{g-1}} = E\left(-\frac{\partial^2 \ln L}{\partial c_g \partial c_{g-1}}\right) = \sum_r \bar{N}_r \sum_j \frac{1}{P_{j,g}(X_r)} \frac{\partial P_{j,g}^*(X_r)}{\partial c_g} \frac{\partial P_{j,g-1}^*(X_r)}{\partial c_{g-1}} \quad (59)$$

or

$$V_{c_g, c_{g+1}} = E\left(-\frac{\partial^2 \ln L}{\partial c_g \partial c_{g+1}}\right) = \sum_r \bar{N}_r \sum_j \frac{1}{P_{j,g+1}(X_r)} \frac{\partial P_{j,g}^*(X_r)}{\partial c_g} \frac{\partial P_{j,g+1}^*(X_r)}{\partial c_{g+1}} \quad (60)$$

and finally, if $|g - g'| \geq 2$

$$V_{c_g c_{g'}} = 0 \quad (61)$$

The first derivative of the cumulative standard normal distribution function with respect to c_k is given by (47).

Applying the Newton-Raphson iterative method to any q-th stage estimate of the category parameters, c_g's, we obtain the matrix (62):

$$\begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \cdot \\ \cdot \\ \hat{c}_g \\ \cdot \\ \cdot \\ \hat{c}_{m-1} \end{bmatrix}_{q+1} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \cdot \\ \cdot \\ \hat{c}_g \\ \cdot \\ \cdot \\ \hat{c}_{m-1} \end{bmatrix}_{q} + \begin{bmatrix} V_{11} & V_{12} & & & & & \\ V_{21} & V_{22} & V_{23} & & & & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & V_{gg} & & & \\ \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \\ 0 & 0 & 0 & 0 & \cdot & V_{m-1,m-1} & \end{bmatrix}^{-1} \begin{bmatrix} t_1 \\ t_2 \\ \cdot \\ \cdot \\ t_g \\ \cdot \\ \cdot \\ t_{m-1} \end{bmatrix}$$

where $V_{gg}=Vc_gc_g$ and $t_g=t_{c_g}$.

Step Parameter Estimation

Since the step parameter d_g, which is the parameter d_k for the specific step $k=g$, is contained in all $P_{jk}(\theta)$ $(k=1,2,\ldots,m)$ as shown in (20), the first derivative of the probability of specific response pattern with respect to d_g is given by

$$\frac{\partial \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{jk}(\theta_i)]^{W_{jki}}}{\partial d_g}$$

$$= \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{jk}(\theta)]^{W_{jki}} \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{W_{jki}}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g} \qquad (63)$$

According to the EM algorithm and Equation (63), the maximum likelihood function (27) with respect to d_g is written as

$$t_{d_g} = \sum_{f=1}^{F} \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{\bar{r}_{jkf}}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g}$$

$$= \sum_{f=1}^{F} \sum_{j=1}^{n} a_j \sum_{k=g}^{m} [\bar{r}_{jkf} - P_{jk}(X_f) \sum_{k=1}^{m} \bar{r}_{jkf}]$$

because

$$\frac{\partial P_{jk}(X_t)}{\partial d_g} = \frac{\partial}{\partial d_g}\left[\frac{\exp[\sum_{v=1}^{k} Z_{jv}(X_t)]}{\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]}\right]$$

$$= \frac{1}{[\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]^2}\left[\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]\frac{\partial[\exp[\sum_{v=1}^{k} Z_{jv}(X_t)]]}{\partial d_g}\right.$$

$$\left. - \exp[\sum_{v=1}^{k} Z_{jv}(X_t)]\frac{\partial[\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_g}\right],$$

$$\sum_{k=1}^{m}\frac{\bar{r}_{jkt}}{P_{jk}(X_t)}\frac{1}{\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]}\frac{\partial[\exp[\sum_{v=1}^{k} Z_{jv}(X_t)]]}{\partial d_g}$$

$$= a_j\sum_{k=g}^{m}\bar{r}_{jkt}$$

and

$$\frac{\bar{r}_{jkt}}{\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]}\frac{\partial[\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_g}$$

$$= a_j\sum_{k=1}^{m}\bar{r}_{jkt}\frac{\sum_{c=g}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]}{\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]}$$

$$= a_j\sum_{k=1}^{m}\bar{r}_{jkt}\sum_{c=g}^{m}P_{jc}(X_t)$$

$$= a_j\sum_{k=g}^{m}P_{jk}(X_t)\sum_{k=1}^{m}\bar{r}_{jkt}$$

The entry of the information matrix, then, becomes

$$V_{d_g d_{g'}} = \sum_{t=1}^{T}\bar{N}_t\sum_{j=1}^{n}\sum_{k=1}^{m}\frac{1}{P_{jk}(X_t)}\frac{\partial P_{jk}(X_t)}{\partial d_g}\frac{\partial P_{jk}(X_t)}{\partial d_{g'}}$$

$$= \sum_{t=1}^{T}\bar{N}_t\sum_{j=1}^{n}a_j^2\left[\sum_{k=g}^{m}P_{jk}(X_t)-\left[\sum_{k=g}^{m}P_{jk}(X_t)\right]\left[\sum_{k=g'}^{m}P_{jk}(X_t)\right]\right]$$

$$= \sum_{t=1}^{T}\bar{N}_t\sum_{j=1}^{n}a_j^2\left[\sum_{k=g}^{m}P_{jk}(X_t)\right]\left[1-\sum_{k=g'}^{m}P_{jk}(X_t)\right] \quad \text{for } g' \le g$$

because

$$\sum_{k=1}^{m}\frac{1}{P_{jk}(X_t)}\frac{\partial P_{jk}(X_t)}{\partial d_g}\frac{\partial P_{jk}(X_t)}{\partial d_{g'}}$$

$$= \sum_{k=1}^{m}\frac{1}{P_{jk}(X_t)}\left[\frac{1}{\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]}\frac{\partial[\exp[\sum_{v=1}^{k} Z_{jv}(X_t)]]}{\partial d_g}\right.$$

$$- \frac{P_{jk}(X_t)}{\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]} \frac{\partial[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_g}] \cdot$$

$$[\frac{1}{\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]} \frac{\partial[\exp[\sum_{v=1}^{k} Z_{jv}(X_t)]]}{\partial d_{g'}}$$

$$- \frac{P_{jk}(X_t)}{\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]} \frac{\partial[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_{g'}}]$$

$$- \sum_{k=1}^{m} \frac{1}{P_{jk}(X_t)} \frac{1}{[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]^2} \frac{\partial[\exp[\sum_{v=1}^{k} Z_{jv}(X_t)]]}{\partial d_g} \cdot$$

$$\frac{\partial[\exp[\sum^{k} Z_{jv}(X_t)]]}{\partial d_{g'}}$$

$$= a_j^2 \sum_{k=g}^{m} P_{jk}(X_t) \quad ,$$

$$\sum_{k=1}^{m} \frac{1}{[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]^2} \frac{\partial[\exp[\sum_{v=1}^{k} Z_{jv}(X_t)]]}{\partial d_g} \cdot$$

$$\frac{\partial[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_{g'}}$$

$$= a_j^2 \sum_{k=g}^{m} P_{jk}(X_t) \sum_{k=g'}^{m} P_{jk}(X_t) \quad ,$$

$$\sum_{k=1}^{m} \frac{1}{[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]^2} \frac{\partial[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_g} \cdot$$

$$\frac{\partial[\exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_{g'}}$$

$$= a_j^2 \sum_{k=g'}^{m} P_{jk}(X_t) \sum_{k=g}^{m} P_{jk}(X_t) \quad ,$$

and

$$\sum_{k=1}^{m} \frac{P_{jk}(X_t)}{[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]^2} \frac{\partial[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_g} \cdot$$

$$\frac{\partial[\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_{jv}(X_t)]]}{\partial d_{g'}}$$

$$= a_j^2 \sum_{k=1}^{m} P_{jk}(X_t) \sum_{k=g}^{m} P_{jk}(X_t) \sum_{k=g'}^{m} P_{jk}(X_t)$$

$$= a_j^2 \sum_{k=g}^{m} P_{jk}(X_r) \sum_{k=g'}^{m} P_{jk}(X_r)$$

Since d_1 is defined 0, the orders of the gradient vector (t_k) and the information matrix V are m-1 and (m-1)x(m-1) respectively. The Newton-Raphson equation is written as

$$\begin{bmatrix} \hat{d}_2 \\ \hat{d}_3 \\ \cdot \\ \hat{d}_g \\ \cdot \\ \hat{d}_m \end{bmatrix}_{q+1} = \begin{bmatrix} \hat{d}_2 \\ \hat{d}_3 \\ \cdot \\ \hat{d}_g \\ \cdot \\ \hat{d}_m \end{bmatrix}_{q} + \begin{bmatrix} V_{22} & & & & & \\ V_{32} & V_{33} & & & & \\ \cdot & \cdot & & & & \\ \cdot & \cdot & V_{gg} & & & \\ \cdot & \cdot & \cdot & & & \\ V_{m2} & V_{m3} & V_{mg} & \cdot & \cdot & V_{mm} \end{bmatrix}^{-1} \begin{bmatrix} t_2 \\ t_3 \\ \cdot \\ \cdot \\ t_g \\ \cdot \\ t_m \end{bmatrix} \qquad (64)$$

During the process of the iteration, the value of $d_1=0$ is used to evaluate the entries of the gradient vector and the information matrix. The information matrix is symmeric. The inversion of the information matrix in (62) and (64) is programmed according to the bordering method (Faddeeva, 1959, pp.105-111).

The MML-EM Algorithm

The MML-EM algorithm consists of two steps. The first is the Expectation step (the E-step) where the provisional expected frequency and the provisional expected sample size are computed by Equations (37) and (45) respectively. Then, in the Maximization step (the M-step), the maximum likelihood equations (46) for each item and (62) or (64) for m-1 categories are solved by the Newton-Raphson iterative method. The numerical procedure of the MML-EM algorithm is outlined as follows:

1. X_f is the value of the independent variable at level f. Define X_f so that $\sum X_f = 0$. Compute the corresponding weights by quadrature or empirical distribution.
2. Define an initial set of $\hat{a}_j$. In the Fortran program, the set of $\hat{a}_j$ is initialized as $\hat{a}_j=1$, $j=1,2,\ldots,n$.
3. Define an initial set of $\hat{b}_j$. In the Fortran program, the set of $\hat{b}_j$ is initialized as $\hat{b}_j=0$, $j=1,2,\ldots,n$.
4. Define an initial set of $\hat{c}_k$ or $\hat{d}_k$. In the Graded Response model, the normal distribution is divided into m equal probabilities and the m-1 corresponding normal deviates are assigned to the initial set of $\hat{c}_k$, $k=1,2,\ldots,m-1$. Therefore, $\sum\hat{c}_k=0$ or $c_.=0$. In the Rating Scale model, all initial values of the step parameters ($\hat{d}_k$) are 0.
5. For provisional $\hat{a}_j$, $\hat{b}_j$, and $\hat{c}_k$ or $\hat{d}_k$, compute $L_l(X_f)$, $f=1,2,\ldots,F$, and $\tilde{P}_l=\sum L_l(X_f)A(X_f)$ for pattern l, $l=1,2,\ldots,s$. Accumulate $\bar{r}_{jkf}$ and $\bar{N}_f$ by summing with respect to l.
6. Improve a set of estimates $\hat{c}_k$ or $\hat{d}_k$ using the Newton-Raphson procedure.
7. Repeat the processes 5 and 6 until the estimates $\hat{c}_k$ or $\hat{d}_k$ become stable to the required number of places. In the Fortran program, the precision is set as, say, $|\hat{c}_k - \hat{c}'_k| < .01$ or $|\hat{d}_k - \hat{d}'_k| < .01$ when $\hat{c}_k$ or $\hat{d}_k$ is the previous estimates, $k=1,2,\ldots,m-1$.

8. Center these estimates at $\hat{c}_\cdot=0$ or $\hat{d}_\cdot=0$.
9. For provisional $\hat{a}_j$, $\hat{b}_j$, and $\hat{c}_k$ or $\hat{d}_k$, compute $L_i(X_r)$, $\tilde{P}_i$, $\bar{r}_{jkr}$, and $\bar{N}_r$.
10. Improve a set of estimates $\hat{a}_j$ and $\hat{b}_j$ for an item using the Newton-Raphson procedure.
11. Apply the processes 9 and 10 for the next item estimates, $\hat{a}_j$ and $\hat{b}_j$. Repeat these processes for all item estimates $\hat{a}_j$ and $\hat{b}_j$, j=1,2,...,n.
12. After the estimation of all item parameters, a_j and b_j, check whether or not all estimates, $\hat{a}_j$ and $\hat{b}_j$ are met by pre-established precision criteria, say, $|\hat{a}_j - \hat{a}'_j| < .01$ and $|\hat{b}_j - \hat{b}'_j| < .01$ for j=1,2,...,n. If not, repeat the processes 9, 10 and 11 until all estimates $\hat{a}_j$ and $\hat{b}_j$ become stable.
13. The processes from 5 to 12 are considered to be a single cycle of iteration. This cycle is repeated until all estimates, $\hat{a}_j$, $\hat{b}_j$, and $\hat{c}_k$ or $\hat{d}_k$ converge simultaneously to the pre-established criteria.

The Analysis of Fit

Once we have completed the estimation of parameters for the specific model, an analysis-of-fit must be done in order to investigate the validity of the assumed model for a specific data set. If the model is not suitable for explaining the behavior of a given data set, the analysis-of-fit can provide valuable information about how and why

the specific model fails to fit the subjects' response patterns.

The chi-square test is the most popular statistic for the purpose of the analysis-of-fit. The chi-square statistics for testing the assumed model are

$$X^2 = \sum_{i=1}^{m^n}(r_i - N\tilde{P}_i)^2/N\tilde{P}_i \qquad (65)$$

The statistics (65) are approximately distributed as a chi-square random variable with m^n-2n-m degrees of freedom since the m^n terms are summated and 2n item and m-1 category parameters have been estimated. The approximation becomes increasingly accurate as the sample size N increases. If m^n is large relative to N, the frequency table will be sparse and X^2 will tend to be unstable numerically because $N\tilde{P}_i$ will be frequently less than 5 (Bhattacharyya & Johnson, 1977, p.425). Bock and Aitkin (1981) suggested the pooling of the frequencies of patterns with small expectations until all expected frequencies equal or exceed 5. The degree of freedom then becomes s_0-2n-m after the pooling where s_0 is the number of pooled frequencies after being ordered according to their corresponding latent trait estimates. This chi-square is called the pooled chi-square, X^2_{pool}.

Another popular fit-statistic is called the likelihood-ratio chi-square, which is given by

$$G^2 = 2[\sum_{l=1}^{s} r_l \ln(r_l/N\tilde{P}_l)] \qquad (66)$$

The likelihood-ratio chi-square ststistic, G^2, has a large sample properties similar to X^2. The degree of freedom of G^2 is the same as X^2. The G^2 is not computable if $r_i=0$. Therefore, G^2 and G^2_{pool} are computed only for the realized response patterns. When m^n is extremely large, the calculation of m^n expected frequencies itself needs a considerably long computational time. In this case, only G^2 and G^2_{pool} are computed.

A X^2 random variable is the sum of many identically distributed independent random variables with finite moments. According to the central limit theorem, such a sum approaches normality as the degree of freedom, n, increases. The direct comparison of X^2 statistics is cumbersome because they depend on the parameter n. Standardization of X^2 statistics makes the comparison possible without computing the probability levels. The mean and the variance of X^2_n are

$$E(X^2_n) = n$$

and

$$Var(X^2_n) = 2n$$

Mosteller and Rourke (1973,p.150) show that the shape of the normal distribution begins to emerge when n is more than 20. A normal approximation of fit-statistics is denoted by $Z(X^2)$ or $Z(G^2)$ in this paper.

Structural Item Response Model

Recently, statistical methods of qualitative data analysis has progressed greatly (Bishop, Fienberg, & Holland, 1975; Bock, 1981). The new qualitative methodologies are structural rather than descriptive. A model with parameters is constructed, and parameters are estimated. Many points of similarity exist between the new qualitative methods and the more familiar qualitative techniques of analysis of regression, analysis of variance, and correlational analysis (Bock, 1981, p.1).

Bock (1975) analyzed qualitative data sets based on the techniques which he developed for univariate and multivariate qualitative data analyses. In analyzing ordered categorical data, Bock (1975, pp.541-546) incorporated the matrix that represents a structure among subject groups with the normal response function. To utilize such a design matrix makes a qualitative analysis of structured data more elegant.

In this paper, the design matrix is incorporated with the Graded Response model (9) to test the significance of group differences in terms of item location parameters (b_j). Since the model represents the multiple-item and multiple-category response behavior of subjects in structured groups, the model is referred to as a structural item response model. The structural item response model developed in this paper is simply the three-parameter normal ogive latent

trait model reparameterized by utilizing the design or model matrix. Therefore, the model can be viewed as an analytic model rather than a measurement model with new parameters.

The structural item response model developed from the Graded Response model is

$$P^{*}_{rjk}(\theta) = \int_{-\infty}^{Z_{rjk}(\theta)} \phi(t)dt$$

$$= \int_{-\infty}^{a_j(\theta - b_{rj} + c_k)} \phi(t)dt \qquad (67)$$

where a_j is the slope parameter for item j (j=1,2,...,h,...,n), b_{rj} is the location parameter for group r (r=1,2,...,u,...,q) and item j, and c_k is the category parameter for category k (k=1,2,...,g,...,m). The only difference between the Graded Response model (9) and the structural item response model (67) is that in the latter the values of location parameter vary among q subject groups as well as among n items.

We assume that the matrix of the item location parameters B can be described by the linear model

$$\underset{qxn}{B} = \underset{qxt}{M}\ \underset{txn}{F^{*}} \qquad (68)$$

where M is the design or model matrix, which is common for all items j=1,2,...,n, and F^{*} is the matrix of effects. If M is of deficient rank $1<t\leq q$, we reparameterize (68) as

$$\underset{qxn}{B} = \underset{qxl}{K}\ \underset{lxt}{L}\ \underset{txn}{F^{*}}$$

$$= \underset{qxl}{K}\ \underset{lxn}{F} \qquad (69)$$

or

$$\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1h} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2h} & \cdots & b_{2n} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ b_{u1} & b_{u2} & \cdots & b_{uh} & \cdots & b_{un} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ b_{q1} & b_{q2} & \cdots & b_{qh} & \cdots & b_{qn} \end{bmatrix}$$

$$= \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1v} & \cdots & K_{1l} \\ K_{21} & K_{22} & \cdots & K_{2v} & \cdots & K_{2l} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ K_{u1} & K_{u2} & \cdots & K_{uv} & \cdots & K_{ul} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ K_{q1} & K_{q2} & \cdots & K_{qv} & \cdots & K_{ql} \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1h} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2h} & \cdots & f_{2n} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ f_{v1} & f_{v2} & \cdots & f_{vh} & \cdots & f_{vn} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ f_{l1} & f_{l2} & \cdots & f_{lh} & \cdots & f_{ln} \end{bmatrix}$$

where **K** is the basis matrix for the design **M**, and **L** is the parameterization matrix.

In Equation (69), the design or model matrix **M** is factorized into the product of the column basis **K** and the row basis **L**, that is

$$\mathbf{M} = \mathbf{K}\,\mathbf{L} \tag{70}$$

where $|\mathbf{K}'\mathbf{K}| \neq 0$. Thus, the deficient-rank model in terms of the tn parameters in $\mathbf{F}^{*}$ is replaced by the full-rank model in terms of ln linear parameteric functions in **F**. Bock (1975) states, "... reparameterization is perhaps the most transparent from the point of view of interpretation and is in general the most economical computationally (p.239)."

The factorization (70) is not unique. Given both **M** and **L**, however, we may compute **K** from

$$\mathbf{K} = \mathbf{M}\,\mathbf{L}'\,(\mathbf{L}\,\mathbf{L}')^{-1} \tag{71}$$

A choice of the model matrix **M** depends on the structure of research design, and data analysts determine the parameterization matrix **L** according to their statistical hypothesis and interpretation about effects. Bock (1975) discusses several alternative types of contrasts in **L** which may facilitate the interpretation of effects **F**.

Estimation

The likelihood for all subject groups is expressed as the product of likelihood (26) such that

$$L = \prod_{r=1}^{q} P((r_{ri})|N_r,(P_{ri}))$$

$$= \prod_{r=1}^{q} (N_r!/\prod_{i=1}^{m^n} r_{ri}!) \prod_{i=1}^{m^n} [P_{ri}((W_{rjk}))]^{r_{ri}} \qquad (72)$$

where r_{ri} is the frequency of the i-th response pattern in group r ($i=1,2,\ldots,m^n$), N_r is the number of subjects in group r. We further obtain the following equations:

$$r_{ri} = N_r p_{ri} \qquad (73)$$

$$\sum_{r=1}^{q}\sum_{i=1}^{m^n} r_{ri} = \sum_{r=1}^{q} N_r = N \qquad (74)$$

$$\sum_{r=1}^{q}\sum_{i=1}^{m^n} p_{ri} = q \qquad (75)$$

and

$$E(p_{ri}) = P_{ri} \qquad (76)$$

The term $P_{ri}((W_{rjk}))$ in Equation (72) can be calcul lated in the following manner:

$$P_{ri}((W_{rjk})) = \int_{-\infty}^{\infty} P_{ri}((W_{rjk})|\theta)\phi(\theta)d\theta \tag{77}$$

$$P_{ri}((W_{rjk})|\theta) = \prod_{j=1}^{n}\prod_{k=1}^{m}[P_{rjk}(\theta)]^{W_{rjki}} \tag{78}$$

$$P_{rjk}(\theta) = P^{*}_{rjk}(\theta) - P^{*}_{rj,k-1}(\theta) \tag{79}$$

$$= \int_{-\infty}^{a_j(\theta-b_{rj}+c_k)} \phi(t)dt - \int_{-\infty}^{a_j(\theta-b_{rj}+c_{k-1})} \phi(t)dt$$

$$= \int_{-\infty}^{a_j(\theta-\sum_{c=1}^{l}K_{rc}f_{cj}+c_k)} \phi(t)dt - \int_{-\infty}^{a_j(\theta-\sum_{c=1}^{l}K_{rc}f_{c}+c\quad)} \phi(t)dt$$

Following the same derivation process described previously, we obtained the following elements of gradient vector and information matrix to estimate the slope parameter a_h for specific item j=h and the effects f_{ch} for j=h and c=1,2,...,v,...,l:

$$t_{a_h} = \sum_{r}^{q}\sum_{f}^{r}\sum_{k}^{m} \frac{\bar{r}_{rhkf}}{P_{rhk}(X_f)} \frac{\partial P_{rhk}(X_f)}{\partial a_h} \tag{80}$$

$$t_{f_{vh}} = \sum_{r}^{q}\sum_{f}^{r}\sum_{k}^{m} \frac{\bar{r}_{rhkf}}{P_{rhk}(X_f)} \frac{\partial P_{rhk}(X_f)}{\partial f_{vh}} \tag{81}$$

$$V_{a_h a_h} = \sum_{r}^{q}\sum_{f}^{r} \bar{N}_{ri} \sum_{k}^{m} \frac{1}{P_{rhk}(X_f)} \left[\frac{\partial P_{rhk}(X_f)}{\partial a_h}\right]^2 \tag{82}$$

$$V_{f_{vh}f_{v'h}} = \sum_{r}^{q}\sum_{f}^{r} \bar{N}_{ri} \sum_{k}^{m} \frac{1}{P_{rhk}(X_f)} \frac{\partial P_{rhk}(X_f)}{\partial f_{vh}} \frac{\partial P_{rhk}(X_f)}{\partial f_{v'h}} \tag{83}$$

and

$$V_{a_h f_{vh}} = \sum_{r}^{q}\sum_{f}^{r} \bar{N}_{ri} \sum_{k}^{m} \frac{1}{P_{rhk}(X_f)} \frac{\partial P_{rhk}(X_f)}{\partial a_h} \frac{\partial P_{rhk}(X_f)}{\partial f_{vh}} \tag{84}$$

where

$$\frac{\partial P_{rhk}(X_f)}{\partial a_h}$$

$$= \phi[Z_{rhk}(X_f)][X_f - \sum_{c=1}^{l} K_{rc} f_{ch} + c_k]$$

$$- \phi[Z_{rh,k-1}(X_f)][X_f - \sum_{c=1}^{l} K_{rc} f_{ch} + c_{k-1}] \quad (85)$$

$$\frac{\partial P_{rhk}(X_f)}{\partial f_{vh}}$$

$$= -a_h K_{rv}[\phi[Z_{rhk}(X_f)] - \phi[Z_{rh,k-1}(X_f)]] \quad (86)$$

and $\phi[Z_{rhk}(X_f)]$ is the standard normal density at the deviate $Z_{rhk}(X_f) = a_h(X_f - \sum K_{rc} f_{ch} + c_k)$. It should be noted that

$$\sum_{c=1}^{l} K_{rc} \hat{f}_{ch} = \hat{b}_{rh} \quad (87)$$

$\bar{N}_{rf}$ and $\bar{r}_{rjkf}$ are called a provisional expected sample size of group r at the quadrature point f and a provisional expected frequency of the positive response in group r to item j and category k at level f respectively. $\bar{N}_{rf}$ and $\bar{r}_{rhkf}$ are computed in the following manner:

$$\bar{N}_{rf} = \sum_{l_r=1}^{S_r} [r_{l_r} L_{l_r} A(X_f) / \widetilde{P}_{l_r}] \quad (88)$$

and

$$\bar{r}_{rhkf} = \sum_{l_r=1}^{S_r} [r_{l_r} L_{l_r} (A_f) W_{hkl_r} / P_{l_r}] \quad (89)$$

where

$$\widetilde{P}_{l_r} = \sum_{f=1}^{F} \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{rjk}(X_f)]^{W_{jkl_r}} A(X_f) \quad (90)$$

$$L_{l_r} = \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{rjk}(X_f)]^{W_{jkl_r}} \quad (91)$$

r_{l_r} is the number of subjects of the l-th response pattern in group r, $W_{jkl_r}=1$ if the l-th response pattern in group r gives a to item j and category k, and $W_{jkl_r}=0$ if the response is incorrect. It should be noted that the l_r-th response pattern is not identical among q groups since some specific response patterns may not appear within certain subject groups.

For each item $j=h$, the Newton-Raphson method at the q-th interative stage is, then, expressed in a matrix form (92):

$$\begin{pmatrix} \hat{a}_h \\ \hat{f}_{1h} \\ \hat{f}_{2h} \\ \vdots \\ \hat{f}_{vh} \\ \vdots \\ \hat{f}_{lh} \end{pmatrix}_{q+1} = \begin{pmatrix} \hat{a}_h \\ \hat{f}_{1h} \\ \hat{f}_{2h} \\ \vdots \\ \hat{f}_{vh} \\ \vdots \\ \hat{f}_{lh} \end{pmatrix}_{q} + \begin{pmatrix} Va_ha_h & & & & \\ Va_hf_{1h} & Vf_{1h}f_{1h} & & & \\ Va_hf_{2h} & Vf_{1h}f_{2h} & & & \\ \vdots & \vdots & & & \\ Va_hf_{vh} & Vf_{1h}f_{vh} & \cdots & Vf_{vh}f_{vh} & \\ \vdots & \vdots & & \vdots & \\ Va_hf_{lh} & Vf_{1h}f_{lh} & \cdots & Vf_{vh}f_{lh} & \cdots & Vf_{lh}f \end{pmatrix}^{-1} \begin{pmatrix} t_{ah} \\ t_{f} \\ t_{f} \\ \vdots \\ t_{f} \\ \vdots \\ t_{f} \end{pmatrix}$$

Computation of the elements in the gradient vector and the information matrix in (92) can be further simplified if we express (80), (81), (82), (83), and (84) in the following manner:

$$t_{ah} = \sum_{r=1}^{q} t_{ah}^{(r)} \tag{93}$$

$$t_{fvh} = \sum_{r=1}^{q} K_{rv} t_{bh}^{(r)} \tag{94}$$

$$Va_ha_h = \sum_{r=1}^{q} Va_ha_h^{(r)} \tag{95}$$

$$Vf_{vh}f_{v'h} = \sum_{r=1}^{q} K_{rv}K_{rv'}Vb_hb_h^{(r)} \qquad (96)$$

and

$$Va_hf_{vh} = \sum_{r=1}^{q} K_{rv}Va_hb_h^{(r)} \qquad (97)$$

The $t_{ah}^{(r)}$, $t_{bh}^{(r)}$, $Vb_hb_h^{(r)}$, and $Va_hb_h^{(r)}$ are the elements in the gradient vector and the information matrix in (46), which are independently computed within group r (r=1,2,...,q) by using (36), (38), (42), (43), and (44). The elements in (92) are, therefore, the summation of these statistics over q groups. For the estimation of the effect f_{vh}, the v-th and v'-th column vectors in **K** are used as weights.

In the same manner, we obtain the following elements of gradient vector and information matrix to estimate the category paramete

$$t_{C_g} = \sum_{rf}^{qf}\sum_{j}^{n}\left[\frac{\bar{r}_{rjgf}}{P_{rjg}(X_f)} - \frac{\bar{r}_{rj,g+1,f}}{P_{rj,g+1}(X_f)}\right]\frac{\partial P^{*}_{rjg}(X_f)}{\partial c_g} \qquad (98)$$

$$Vc_gc_g = \sum_{rf}^{qf}\bar{N}_{rf}\sum_{j}^{n}\left[\frac{1}{P_{rjg}(X_f)} + \frac{1}{P_{rj,g+1}(X_f)}\right]\left[\frac{\partial P^{*}_{rjg}(X_f)}{\partial c_g}\right]^2 \qquad (99)$$

$$Vc_gc_{g-1} = \sum_{rf}^{qf}\bar{N}_{rf}\sum_{j}^{n}\frac{1}{P_{rjg}(X_f)}\,\frac{\partial P_{rjg}^{*}(X_f)}{\partial c_g}\,\frac{\partial P^{*}_{rj,g+1}(X_f)}{\partial c_g} \qquad (100)$$

$$Vc_gc_{g'} = 0 \text{ if } |g-g'|\geq 2 \qquad (101)$$

where

$$\frac{\partial P^{*}_{rjg}(X_f)}{\partial c_g} = a_i\,\phi\left[Z_{rjg}(X_f)\right] \qquad (102)$$

The Newton-Raphson method at the q-th iterative stage is, then, expressed as the following equation (103):

$$\begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \cdot \\ \cdot \\ \hat{c}_g \\ \cdot \\ \cdot \\ \hat{c}_{m-1} \end{pmatrix}_{q+1} = \begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \cdot \\ \cdot \\ \hat{c}_g \\ \cdot \\ \cdot \\ \hat{c}_{m-1} \end{pmatrix}_{q} + \begin{pmatrix} V_{11} & V_{12} & & & & & \\ V_{21} & V_{22} & V_{23} & & & & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & V_{gg} & & & \\ \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \\ 0 & 0 & 0 & 0 & \cdot & V_{m-1,m-1} & \end{pmatrix}^{-1} \begin{pmatrix} t_1 \\ t_2 \\ \cdot \\ \cdot \\ t_g \\ \cdot \\ \cdot \\ t_{m-1} \end{pmatrix}$$

where $V_{gg}=Vc_gc_g$ and $t_g=t_{c_g}$.

Computation of the elements in the gradient vector and the information matrix in (103) can be further simplified if we express (98), (99), and (101) in the following manner:

$$t'_{cg} = \sum_{r=1}^{q} t_{cg}^{(r)} \tag{104}$$

$$Vc_gc_g = \sum_{r=1}^{q} Vc_gc_g^{(r)} \tag{105}$$

and

$$Vc_gc_{g-1} = \sum_{r=1}^{q} Vc_gc_{g-1}^{(r)} \tag{106}$$

The $t_{cg}^{(r)}$, $Vc_gc_g^{(r)}$, and $Vc_gc_{g-1}^{(r)}$ are the elements in the gradient vector and the information matrix in (62), which are independently computed within group r (r=1,2,...,q) by using (57), (58), and (59). The elements in (103) are, therefore, the summation of these statistics over q groups.

The MML-EM Algorithm

The step-by-step procedure of the MML-EM algorithm for the structural item response model is outlined as follows:

1. Construct the model matrix M according to a research design and the reparameterization matrix L according to statistical hypotheses.
2. X_f is the value of the independent variable at level f. Define X_f so that $\sum X_f=0$. Compute the corresponding weights by quadrature or empirical distribution.
3. Define an initial set of $\hat{a}_j$. In the Fortran program, the set of $\hat{a}_j$ is initialized as $\hat{a}_j=1$, j=1,2,...,n.
4. Define an initial set of $\hat{f}_{cj}$. In the Fortran program, the set of $\hat{f}_{cj}$ is initialized as $\hat{f}_{cj}=0$ where c=1,2,...1 and j=1,2,...,n.
5. Define an initial set of $\hat{c}_k$. In the Fortran program, the normal distribution is divided into m equal probabilities and the m-1 corresponding normal deviates are assigned to the initial set of $\hat{c}_k$, k=1,2,...,m-1.
6. For provisional $\hat{a}_j$, $\hat{f}_{cj}$, and $\hat{c}_k$ for j=1,2,...,n, c=1,2,...,1, and k=1,2,...,m-1, compute $L_{l_r}(X_f)$ for f=1,2,...,F, and $\tilde{P}_{l_r}=\sum L_{l_r}A(X_f)$ for pattern $l_r=1,2,...,s_r$. Accumulate $\bar{r}_{rjkf}$ and $\bar{N}_{rf}$ by summing with respect to l_r ($l_r=1,2,...,s_r$).
7. Compute $t_{ck}^{(r)}$, $Vc_kc_k^{(r)}$, and $Vc_kc_{k-1}^{(r)}$ for k=1,2,...,m-1.

8. Repeat the processes 6 and 7 for r=1,2,...,q. The t_{ck}, Vc_kc_k, and Vc_kc_{k-1} can be computed by using $t_{ck}^{(r)}$, $Vc_kc_k^{(r)}$, and $Vc_kc_{k-1}^{(r)}$ with respect to r (r=1,2,...,q).
9. Improve a set of estimates $\hat{c}_k$ by using the Newton-Raphson method (103).
10. Repeat the processes 6, 7, 8, and 9 until the estimates $\hat{c}_k$ (k=1,2,...,m-1) become stable to the required number of places. In the Fortran program, the precision is set as, say, $|\hat{c}_k-\hat{c}'_k|<.01$ when $\hat{c}_k$ is the previous estimates, k=1,2,...,m-1.
11. Center these estimates at $\hat{c}_.=0$.
12. For provisional $\hat{a}_j$, $\hat{f}_{cj}$, and $\hat{c}_k$ for j=1,2,...,n, c=1,2,...,1, and k=1,2,...,m-1, compute $L_{l_r}(X_r)$ for f=1,2,...,F and l_r=1,2,...,s_r and $\widetilde{P}_{l_r}=\sum L_{l_r}A(X_r)$ fpr l_r=1,2,...,s_r.
13. Accumulate $\bar{r}_{rjk_r}$ and $\bar{N}_{rr}$ by summing with respect to l_r (l_r=1,2,...,s_r).
14. Compute $t_{aj}^{(r)}$, $Va_ja_j^{(r)}$, $Vb_jb_j^{(r)}$, and Va_jb_j for an item by utilizing the row vector in K.
15. Repeat the processes 12, 13, and 14 for r=1,2,...,q. The t_{aj}, t_{fcj}, Va_ja_j, $Vf_{cj}f_c'_j$ and Va_jf_{cj} for an item j and c=1,2,...,1 can be computed by summing $t_{aj}^{(r)}$, $Va_ja_j^{(r)}$, $Vb_jb_j^{(r)}$, and $Va_jb_j^{(r)}$ with respect to r (r=1,2,...,q). The column vectors in K are used as weights.

16. Improve an estimate $\hat{a}_j$ and a set of estimates $\hat{f}_{cj}$ for an item j and c=1,2,...,l using the Newton-Raphson method (92).

17. Apply the processes 12, 13, 14, 15, and 16 for the next item estimates, $\hat{a}_j$ and $\hat{f}_{cj}$ (c=1,2,...,l). Repeat these processes for all item estimates $\hat{a}_j$ and $\hat{f}_{cj}$'s, j=1,2,...,n.

18. After the estimation of all item parameters, a_j and f_{cj}, check whether or not all estimates, $\hat{a}_j$ and $\hat{f}_{cj}$ are met by pre-established precision criteria, say, $|\hat{a}_j-\hat{a}'_j|<.01$ and $|\hat{f}_{cj}-\hat{f}'_{cj}|<.01$ for j=1,2,...,n and c=1,2,...,l. If not, repeat the processes 12, 13, 14, 15, 16, and 17 until all estimates $\hat{a}_j$ and $\hat{f}_{cj}$ become stable.

19. The processes from 6 to 18 are considered to be a single cycle of iteration. This cycle is repeated until all estimates, $\hat{a}_j$, $\hat{f}_{cj}$, and $\hat{c}_k$ converge simultaneously to the established criteria.

The MML-EM algorithm for the structural item response model is basically same as that for the Graded Response model. The elements of the gradient vector and the information matrix are computed within each group by utilizing the row vectors in the design basis K. Then, these elements are summed with respect to r (r=1,2,...,q) to compute the elements for the Newton-Raphson method (92) and (103). At this point, the column vectors in K are used as weights.

The Design Basis K for the Two-Way 2 X 2 Crossed Design

In this study, the data from the General Social Surveys are analyzed by the two-way 2 x 2 crossed design. In this section, the design basis K_{AB} for the crossed design will be generated.

The design basis $K_{12\ldots f}$ for the f-way crossed design can be generated on the basis of the one-way design K according to the rule of Kronecker product (Bock, 1975, pp.273-275). Let us denote M as the model matrix for the one-way design with two levels, that is

$$\underset{2\times 1}{\underline{b}_j} = \underset{2\times 3}{M}\ \underset{3\times 1}{\underline{f}^*_j} \tag{107}$$

or

$$\begin{pmatrix} b_{1j} \\ b_{2j} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} f^*_{0j} \\ f^*_{1j} \\ f^*_{2j} \end{pmatrix}$$

where $\underline{b}_j$ and $\underline{f}_j$ are column vectors of B and F^* in (68) respectively. The reparameterization matrix L is chosen as

$$\underset{2\times 1}{\underline{f}_j} = \underset{2\times 3}{L}\ \underset{3\times 1}{\underline{f}^*_j} \tag{108}$$

or

$$\begin{pmatrix} f_{1j} \\ f_{2j} \end{pmatrix} = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} f^*_{0j} \\ f^*_{1j} \\ f^*_{2j} \end{pmatrix}$$

$$= \begin{pmatrix} f^*_{0j} + 1/2(f^*_{1j} + f^*_{2j}) \\ f^*_{1j} - f^*_{2j} \end{pmatrix}$$

We can, then obtain the design basis K such that

$$\underset{2\times 3}{M} = \underset{2\times 2}{K}\ \underset{2\times 3}{L} \tag{109}$$

or

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1/2 \\ 1 & -1/2 \end{pmatrix} \begin{pmatrix} 1 & 1/2 & 1/2 \\ 0 & 1 & -1 \end{pmatrix}$$

The design basis for the two-way 2 x 2 crossed design $K_{A,B}$ can be generated by the Kronecker product of the one-way design basis K in (109), that is

$$\underset{4x4}{K_{A,B}} = \underset{2x2}{K_A} \times \underset{2x2}{K_B} \qquad (110)$$

or

$$\begin{pmatrix} 1 & 1/2 & 1/2 & 1/4 \\ 1 & 1/2 & -1/2 & -1/4 \\ 1 & -1/2 & 1/2 & -1/4 \\ 1 & -1/2 & -1/2 & 1/4 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1/2 \\ 1 & -1/2 \end{pmatrix} \times \begin{pmatrix} 1 & 1/2 \\ 1 & -1/2 \end{pmatrix}$$

The model (110) is a rank-4 model, whose effects represent both the differences between two levels of the main factors A and B and their interaction. Excluding the interaction term in (110) we obtain a rank-3 model. Fitting successively rank-4 and rank-3 models to observation data, we can obtain the residual X^2's or G^2's. These fit-statistics provide tests for the significance of specific model.

Data Analyses

In this research, three data sets were analyzed. All three data sets were extracted from the data file, the General Social Surveys, collected by National Opinion Research Center. The General Social Surveys has been conducted during 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1980, and 1982. The questionnaire form used during those years has

been quite consistent, although a few questions were added or deleted in certain years. Two sets of Likert-type questions were sampled from items in the questionnaire. The first set of items are presented in Table 1, and the second set of items are presented in Table 4. The first set of items, so-called Country, has been used during 1974, 1975, 1977, and 1982. The data set of 1982 (Country '82) is analyzed in Data Analysis 1. The second set of items, so called Satisfaction, has been used during 1973, 1974, 1975, 1976, 1977, 1978, 1980, and 1982. In Data Analyses 2 and 3, the data sets of 1974 and 1982 are analyzed (Satisfaction '74 and '82).

Original Likert-scales employed for Country and Satisfaction have 10 points and 7 points respectively. The original scales are modified to facilitate the convergence rate. The modified Likert-scales for Country and Satisfaction are presented in Tables 1 and 4 respectively. In addition, the precision of convergence was set at 0.5 in Study 1, and 0.1 in Studies 2 and 3, to ensure the convergence within 20 cycles.

There were 1484 completed interviewed in 1974, and 1506 in 1982. In the Studies 1 and 2, the first 500 cases were sampled from each data set. The response vectors in the original data file does not seem to be in any particular order. Therefore, the 500 sampled subjects can be assumed to correctly represent all respondents in the original

file. In Study 3, we obtain four sub-groups, that is, Group 1 (Male-'74), Group 2 (Male-'82), Group 3 (Female-'74), and Group 4 (Female-'82). The first 300 subjects were sampled from the original data sets. Therefore, totally 1200 subjects were used for the anlysis. For all studies, the distribution of the latent trait variable was assumed to be standard normal. Ten quadrature points were used for the integration.

Data Analysis 1

The original Likert-scale was shortened to a five-point scale. The number of items is 8. Therefore, there are 390,625 possible response patterns. Since 167 response patterns are realized for 500 subjects, the ratio is quite small (0.0004). Nevertheless, the MML-EM algorithm could produce meaningful estimates, shown in Table 3. Comparing the estimates in Table 3 with frequency distributions in Table 2 leads us to believe the estimates are plausible. The problem arised in computing the fit-statistics. Since the ratio of realized patterns is quite small, G^2 or G^2_{pool} tends to be extremely large and the degree of freedom of G^2_{pool} becomes small. On the other hand, the computation of the expected frequencies for more than 300,000 possible response patterns is practically impossible. The fit statistics need to be studied. In Table 3, only G^2_{pool} is presented.

Except the scale constant 1.7 and the sign of item estimates, the MML-EM item estimates based on the Graded Response model and the Rating Scale model are very close. The small range of slope estimates indicates that the quality of the eight items is relatively homogeneous. However, the location estimates vary to a great extent. Russia is the country more likely to be disliked. Canada and England are more likely to be liked. Other countries, Japan, Israel, Brazil, Egypt, and China evoke no particular preference. They are clustered within a small range on the location scale. In this way, the item response model can provide visual representation of the data structure, which is not easily identified from the raw frequencies in Table 2. Category estimates indicate that the categories of "dislike" are less frequently chosen than the categories of "like". These estimates are reasonable since about 78 percent of all respondents answer items by choosing Categories 1, 2, or 3, as shown in Table 2. Step estimates display the same categorical response behavior. The probability of transition from Category 1 to Category 2 is governed by step estimate 1, and so on. Higher estimate value implies easier transition between adjacent categories. The highest step estimate 1 (1.55) indicates that respondents who choose Category 1 are also likely to endorse Category 2. On the other hand, the lowest step estimates (-1.60) indicate that the respondents who choose Category 3 are unlikely to endorce

Category 4. Therefore, most of respondents likely choose Category 1, 2, or 3. Although these step estimates are matched with the categorical estimate values, the interpretation of the step estimates becomes more difficult as the number of estimates increases.

Data Analysis 2

The item response model is suitable for cross-sectional study. The set of items, Satisfaction, was used repeatedly from 1973 to 1982 to different groups of respondents. Since the estimates are sample independent, it is possible to compare between two temporal points, 1974 and 1982, to investigate whether or not people's degree of satisfaction is changed for particular areas.

Category estimates indicate that categories of "less satisfaction" are seldom chosen. Therefore, the location estimates for both models are clustered in a certain direction. All location estimates in the Graded Response model are negative, and all location estimates in the Rating Scale model are positive. In Figure 2, thus, the direction of the location scale is reversed each other to make the figure more comprehensible.

Item estimates in both models show a certain change, but the category or step estimates do not change between 1974 and 1982. The stability of category or step estimates makes the interpretation of item estimates more meaningful.

In the Graded Response model, the satisfaction from City is slightly diminished while the satisfaction from Family is enhanced. Otherwise, the order of the areas, from which people get satisfaction, does not change between those years.

A problem arises when we compare the estimates derived from the Graded Response model with the estimates derived from the Rating Scale model. Since the same data sets were analyzed based on both models, the estimates must be closely matched. However, Table 7 shows some changes in the order of both sets of estimates. This problem may be clarified if we can devise a method to compute the standard error of estimates.

A comparison of the estimates between the Rating Scale model and the Rasch Rating Scale model seems to provide some interesting topics. For the data of 1974, the slope estimate of Item C (Family) is 1.62. The Rasch model assumes that all slope estimates are equally 1. Due to this restriction of the model, the order of location estimates of Item C and Item E between the Rating Scale model and the Rasch Rating Scale model seems to have changed. In general, the polychotomous response model for multiple categorical data will more likely produce varied slope estimates. In such cases, the Rating Scale model may produce a set of different ordered estimates. In addition, Table 7 shows that the Rasch Rating Scale model produces slightly higher fit-statistics, G^2 and G^2_{pool}. This may suggest that a desira-

ble approach is to first fit more general models to the given data and then to decide whether or not simpler models are applicable by checking their slope estimates.

Data Analysis 3

In Data Analysis 2, two data sets, '74 and '82, were analyzed independenly. In Data Analysis 3, both data sets are analyzed simultaneously by incorporating the design matrix M with the Graded Response model. In addition, the sex factor is introduced for this analysis. The two-way design basis $K_{A,}$ shown in (110) is originally prepared, in which the sex (Male-Female) is factor A and the year ('74-'82) is factor B. Each cell contains 300 subjects sampled from two original data sets - Satisfaction '74 and '82. The frequency distributions for classified subjects is presented in Tables 8, 9, 10, and 11.

Two separate data analyses were conducted. In the first analysis, the original two-way design model with the interaction term (Model 1) was fitted to the data . By deleting the interaction term from the original design basis $K_{A,}$ (110), we fitted the rank-3 main-class model (Model 2) to the data in the second analysis. The estimates obtained from these analyses are presented in Tables 12 and 13 respectively. For both analyses, the initial values of effect (location) estimates were computed from the results of Data Analysis 2, which are shown below:

$$\begin{bmatrix} -0.97 & -1.10 & -1.86 & -1.58 & -1.44 \\ -0.82 & -1.27 & -2.25 & -1.68 & -1.61 \\ -0.97 & -1.10 & -1.86 & -1.58 & -1.44 \\ -0.82 & -1.27 & -2.25 & -1.68 & -1.61 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.5 & 0.5 & 0.25 \\ 1 & 0.5 & -0.5 & -0.25 \\ 1 & -0.5 & 0.5 & -0.25 \\ 1 & -0.5 & -0.5 & 0.25 \end{bmatrix} \begin{bmatrix} -0.895 & -1.185 & -2.055 & -1.630 & -1.525 \\ 0 & 0 & 0 & 0 & 0 \\ -0.150 & 0.170 & 0.390 & 0.100 & 0.170 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In both analyses, the iteration process converges after two cycles.

In Tables 12 and 13, the negative value of the location estimates indicates more satisfaction. In addition to the observations discussed in Data Analysis 2, we can notice the following points from the results shown in Tables 12 and 13:

1. People get the least satisfaction from the city in which they live. The satisfaction from City is diminished between '74 and '82. Women are slightly more satisfied with City than men are.
2. People's satisfaction from their own hobby does not change between '74 and '82. Men receive more satisfaction from their hobby than women.
3. People, both men and women, get the most satisfaction from their own family over the years. The lowest slope estimate of Family indicates that family is the area from which people the most commonly receive their satisfaction.
4. On the contrary, the highest slope estimate of Friend may suggest that individual's satisfaction from their friendship varies the most. In other words, it may

suggest that a certain people receive great satisfaction from their friendship, but not so other people. In addition, the interaction effect of Friend seems to be meaningful. While the satisfaction of men from their friendship seems to be diminished between those years, women receive more satisfaction from their friendship in '82 than in '74.

5. Men show more satisfaction from their health than the women do. This tendency has not changed over the years.

The observations discussed above seem quite reasonable. People have become less and less satisfied with the area in which they live. Family is the most important element of people's life, which provides great satisfaction to both men and women. Men seem to enjoy their own hobby, but women still do not seem to have hobbies they are able to enjoy. Women seem to seek their friendship more outside of their family, and they enjoy it. However, the friendship has become less enjoyable for men over the years. Since more men still have to work outside of home than women, they seem to be more conscious of their health than women.

The structural item response model can provide clear image about the structure of the whole data set, which is not easily detectable from raw frequencies. The study about the standard error of estimates derived from the MML-EM algorithm is currently progressed (Louis, 1982). When the

standard error of estimates is available, we can test the significant level of each effect estimate for individual items. At that point, the strutural item response model may become quite valuable tool for analyzing the qualitative data of multiple item responses.

Except Friend, the interaction effect does not seem to be meaningful. The fit-statistics from Model 1 and Model 2 are not different. However, fittting only the main-class model to the data set may lead us to overlook the interesting interaction effects of certain items. Therefore, for a multiple-way crossed research design I recommend that at least the two-way interaction terms should be included in the design basis in the first analysis unless researchers initially have strong reasons to believe that the interaction effects are negligible or conceptually not interpretable.

TABLE 1

Data Analysis 1
Question, Items, and Likert-Scale

Question

You will notice that the boxes on this card go from the highest position of "plus 5" for a country which you like very much, to the lowest postion of "minus 5" for a country you dislike very much. How far up the scale or how far down the scale would you rate the following countries?

Items

A. Russia
B. Japan
C. England
D. Canada
E. Brazil
F. China
G. Israel
H. Egypt

Likert-Scale

Original Scale	+5	+4	+3	+2	+1	-1	-2	-3	-4	-5
Modified Scale		1		2		3		4		5

TABLE 2

Data Analysis 1 Country '82
Frequency Table for Items and Categories

Item	Category	1	2	3	4	5
	A. Russia	14	45	128	83	230
		14	59	187	270	500
	B. Japan	91	169	146	45	49
		91	260	406	451	500
	C. England	189	191	87	16	17
		189	380	467	483	500
	D. Canada	315	118	49	5	13
		315	433	482	487	500
	E. Brazil	68	142	227	42	21
		68	210	437	479	500
	F. China	40	130	169	90	71
		40	170	339	429	500
	G. Israel	104	122	164	60	50
		104	226	390	450	500
	H. Egypt	63	152	201	40	44
		63	215	416	456	500
	Total	884	1069	1171	381	495
		884	1953	3124	3505	4000

TABLE 3

Data Analysis 1 Country '82
Parameter Estimates

	Graded Response Model	Rating Scale Model
Slope		
A. Russia	0.80	0.45
B. Japan	1.07	0.72
C. England	1.03	0.65
D. Canada	0.92	0.64
E. Brazil	1.17	0.88
F. China	1.15	0.74
G. Israel	0.93	0.55
H. Egypt	1.32	1.13
Location		
A. Russia	1.22	-1.81
B. Japan	-0.26	0.41
C. England	-0.98	1.53
D. Canada	-1.68	2.53
E. Brazil	-0.21	0.34
F. China	0.18	-0.22
G. Israel	-0.22	0.36
H. Egypt	-0.14	0.20
Category Step		
1	-1.38	1.55
2	-0.41	0.60
3	0.64	-1.60
4	1.16	-0.55
Cycle	3	5
G^2_{pool} (D.F.)	3249.49 (-15)	3335.17 (-17)

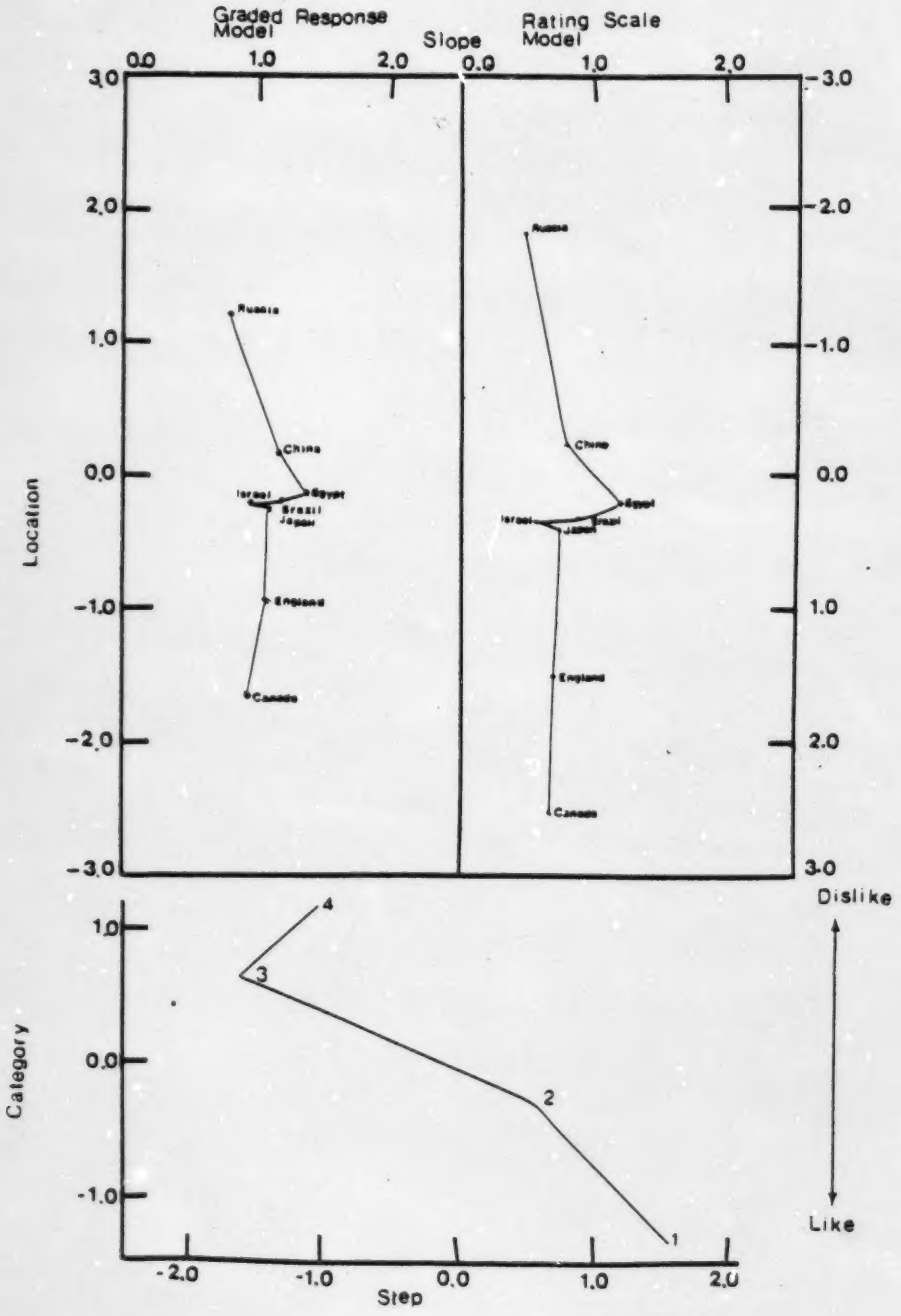

FIGURE 1 Data Analysis 1 Country '82

TABLE 4

Data Analysis.2
Question, Items, and Likert-Scale

Question

For each area of life I am going to name, tell me the number that shows how much satisfaction you get from that area.

Items

A. The city or place you live in.
B. Your non-working activities -- hobbies and so on.
C. Your family life.
D. Your friendships.
E. Your health and physical condition.

Likert-Scale

Original Scale	Modified Scale
A very great deal	1
A great deal	
Quite a bit	2
A fair amount	
Some	3
A little	4
None	

TABLE 5

Data Analysis 2 Satisfaction '74
Frequency Table for Items and Categories

Item	Category	1	2	3	4
	A. City	239	194	37	30
		239	433	470	500
	B. Hobby	272	145	24	59
		272	417	441	500
	C. Family	367	97	9	27
		367	464	473	500
	D. Friend	341	127	17	15
		341	468	485	500
	E. Health	301	137	27	35
		301	438	465	500
	Total	1520	700	114	166
		1520	2220	2334	2500

TABLE 6

Data Analysis 2 Satisfaction '82
Frequency Table for Items and Categories

Item	Category	1	2	3	4
	A. City	191	223	42	44
		191	414	456	500
	B. Hobby	269	148	34	49
		269	417	451	500
	C. Family	374	90	14	22
		374	464	478	500
	D. Friend	329	138	15	18
		329	467	482	500
	E. Health	305	138	24	33
		305	443	467	500
	Total	1468	737	129	166
		1468	2205	2334	2500

TABLE 7

Data Analysis 2 Satisfaction '74 and '82 59
Parameter Estimates

	Graded Response Model		Rating Scale Model		Rasch Rating Scale Model	
	'74	'82	'74	'82	'74	'82
Slope						
A. City	1.03	0.86	0.65	0.50	1.00	1.00
B. Hobby	0.83	0.72	1.02	0.74	1.00	1.00
C. Family	0.86	0.68	1.62	1.09	1.00	1.00
D. Friend	1.14	0.99	1.11	1.13	1.00	1.00
E. Health	0.75	0.71	0.69	0.74	1.00	1.00
Location						
A. City	-0.97	-0.82	1.28	1.14	1.10	0.98
B. Hobby	-1.10	-1.27	0.96	1.17	0.93	1.17
C. Family	-1.86	-2.25	1.42	1.74	1.55	1.92
D. Friend	-1.58	-1.68	1.58	1.52	1.59	1.65
E. Health	-1.44	-1.61	1.54	1.54	1.36	1.53
Category Step						
1	-1.03	-1.18	0.54	0.60	0.55	0.64
2	0.31	0.34	-1.19	-1.28	-1.12	-1.05
3	0.72	0.83	0.65	0.67	0.56	0.40
Cycle	3	4	3	3	3	3
G^2	539.24	557.20	557.88	571.90	606.28	624.81
D.F.	(142)	(149)	(142)	(149)	(147)	(154)
$Z(G^2)$	23.57	23.64	24.67	24.49	26.79	26.82
G^2_{pool}	393.26	433.73	436.94	481.39	501.59	489.47
D.F.	(16)	(16)	(18)	(22)	(24)	(22)
$Z(G^2_{pool})$	66.69	73.84	69.82	64.26	68.93	70.47

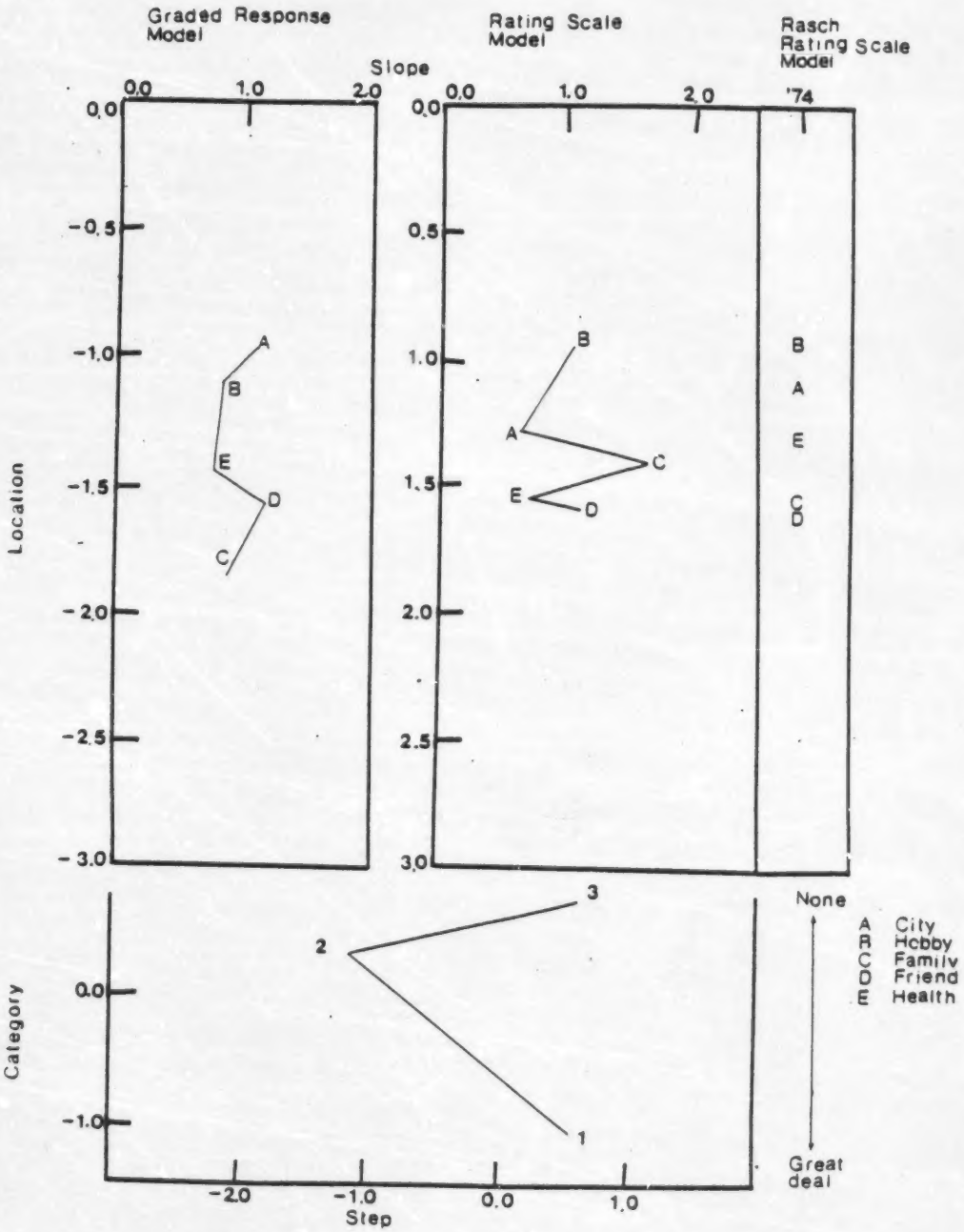

FIGURE 2 Data Analysis 2 Satisfaction '74

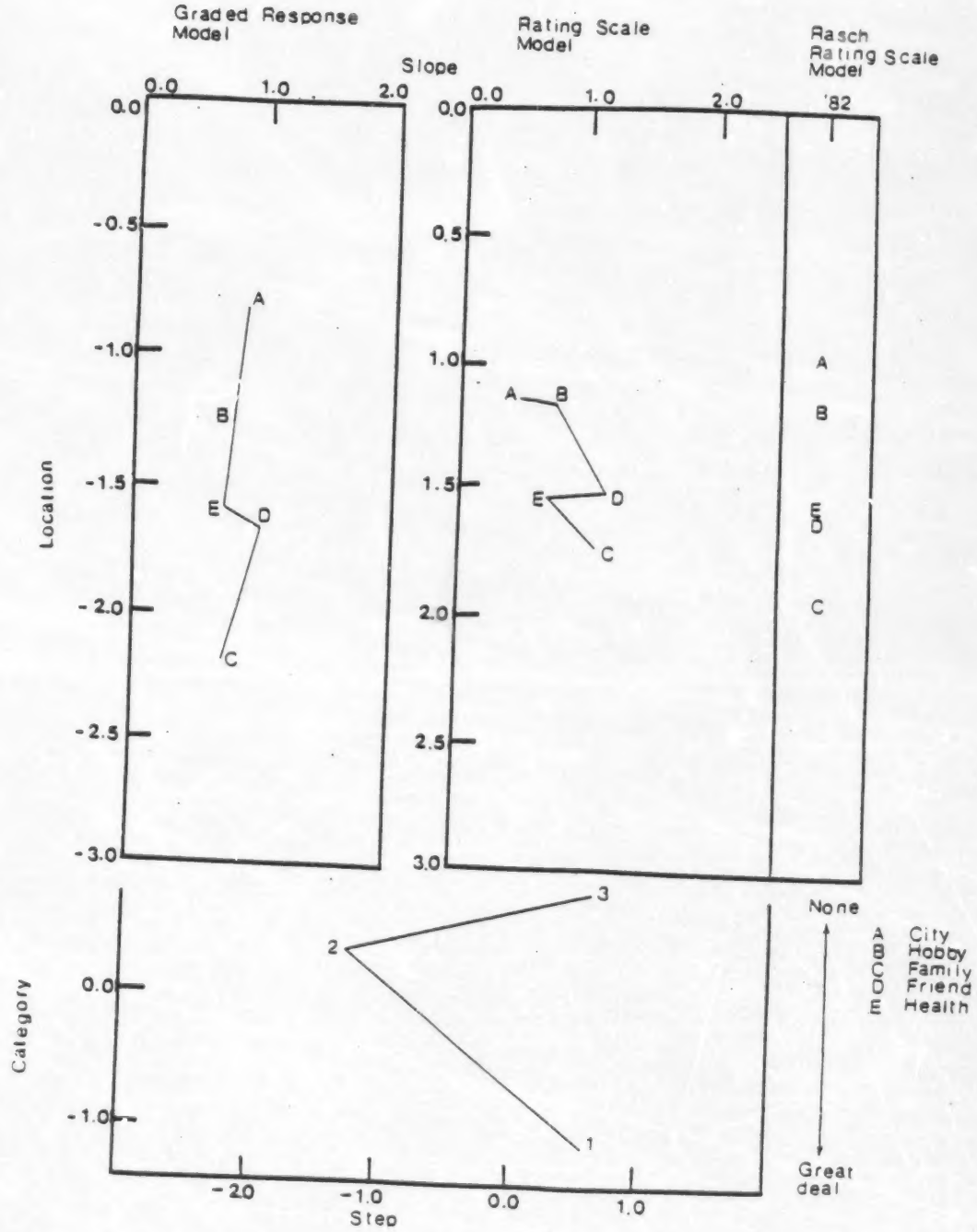

FIGURE 3 Data Analysis 2 Satisfaction '82

TABLE 8

Data Analysis 3 Satisfaction Male-'74
Frequency Table for Items and Categories

ITEM	Category	1	2	3	4
	A. City	142	116	25	17
		142	258	283	300
	B. Hobby	176	82	19	23
		176	258	277	300
	C. Family	223	55	7	15
		223	278	285	300
	D. Friend	216	69	8	7
		216	285	293	300
	E. Health	189	84	12	15
		189	273	285	300
	Total	946	406	71	77
		946	1352	1423	1500

TABLE 9

Data Analysis 3 Satisfaction Male-'82
Frequency Table for Items and Categories

ITEM	Category	1	2	3	4
	A. City	112	139	25	24
		112	251	276	300
	B. Hobby	169	91	18	22
		169	260	278	300
	C. Family	230	41	15	14
		230	271	286	300
	D. Friend	190	89	12	9
		190	279	291	300
	E. Health	195	71	17	17
		195	266	283	300
	Total	896	431	87	86
		896	1327	1414	1500

TABLE 10

Data Analysis 3 Satisfaction Female-'74
Frequency Table for Items and Categories

ITEM	Category	1	2	3	4
	A. City	148	118	13	21
		148	226	279	300
	B. Hobby	158	91	9	42
		158	249	258	300
	C. Family	227	54	5	14
		227	281	286	300
	D. Friend	205	75	10	10
		205	280	290	300
	E. Health	180	81	17	22
		180	261	278	300
	Total	918	419	54	109
		918	1337	1391	1500

TABLE 11

Data Analysis 3 Satisfaction Female-'82
Frequency Table for Items and Categories

ITEM	Category	1	2	3	4
	A. City	127	125	22	26
		127	252	274	300
	B. Hobby	159	84	24	33
		159	243	267	300
	C. Family	228	56	5	11
		228	284	289	300
	D. Friend	210	70	8	12
		210	280	288	300
	E. Health	176	86	16	22
		176	262	278	300
	Total	900	421	75	104
		900	1321	1396	1500

TABLE 12

Data Analysis 3 Parameter Estimates and Fit-Statistics
Model 1: The Two-Way 2 x 2 Crossed Design with Interaction

Item	A.City	B.Hobby	C.Family	D.Friend	E.Health
Slope	0.89	0.73	0.70	1.01	0.71
Location					
Male-74	-1.08	-1.45	-2.20	-1.90	-1.74
Male-82	-0.84	-1.42	-2.28	-1.61	-1.78
Female-74	-1.18	-1.18	-2.31	-1.76	-1.56
Female-82	-0.96	-1.21	-2.35	-1.81	-1.54
Effect					
Constant	-1.01	-1.32	-2.29	-1.77	-1.65
Male-Female	0.10	-0.24	0.08	0.03	-0.21
74-82	-0.23	0.00	0.05	-0.11	0.00
Interaction	-0.01	-0.05	0.03	-0.33	0.06
Category	1	2	3		
	-1.16	0.34	0.82		
	Male-74	Male-82	Female-74	Female-82	Total
G^2	381.01	459.92	376.70	424.04	1641.67
(DF)	(79)	(82)	(81)	(82)	(411)
$Z(G^2)$	24.03	29.51	23.23	26.71	42.92
G^2_{pool}	289.40	325.30	284.58	349.55	1248.83
(DF)	(-10)	(-10)	(-10)	(-9)	(48)
$Z(G^2_{pool})$					122.56

TABLE 13

Data Analysis 3 Parameter Estimates and Fit-Statistics
Model 2: The Two-Way 2 x 2 Crossed Design without Interaction

Item	A.City	B.Hobby	C.Family	D.Friend	E.Health
Slope	0.89	0.73	0.70	1.01	0.71
Location					
Male-74	-1.08	-1.43	-2.21	-1.81	-1.76
Male-82	-0.85	-1.44	-2.27	-1.69	-1.77
Female-74	-1.18	-1.19	-2.30	-1.85	-1.54
Female-82	-0.95	-1.20	-2.36	-1.73	-1.55
Effect					
Constant	-1.01	-1.32	-2.29	-1.77	-1.65
Male-Female	0.10	-0.24	0.08	0.03	-0.21
74-82	-0.22	0.00	0.05	-0.19	0.00
Category	1	2	3		
	-1.16	0.34	0.82		
	Male-74	Male-82	Female-74	Female-82	Total
G^2	382.66	460.71	377.29	425.12	1645.78
DF	(84)	(87)	(86)	(87)	(416)
$Z(G^2)$	23.04	28.33	22.21	25.63	42.64
G^2_{pool}	287.35	329.85	287.60	341.49	1246.29
DF	(-5)	(-5)	(-4)	(-6)	(52)
$Z(G^2_{pool})$					117.11

REFERENCES

Andrich, D. A. A rating formulation for ordered response categories. Psychometrika, 1978, 561-573.

Bhattacharyya, G. K. and Johnson, R. D. Statistical Concepts and Methods. New York: John Wiley & Sons, 1977.

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. Lord and M. Novick, Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Bishop, Y. M. Fienberg, S. E., and Holland, P. W. Discrete Multivariate Analysis: Theory and Practice. Mass.: The MIT Press, 1975.

Bock, R. D. Estimating item parameters and latent trait ability when responses are socred in two or more nominal categories. Psychometrika, 1972, 37, 29-51.

Bock, R. D. Multivariate Statisticsl Methods in Behavioral Research. New York: McGraw-Hill Book Company, 1975.

Bock, R. D. Multivariate Analysis of Qualitative Data. Chicago: National Educational Resources, Inc., 1981.

Bock, R. D. and Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 1981, 41, 443-459.

Dempster, A. P., Laird, N. M. and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. The Journal of the Royal Statistical Society, Series B, 1977, 39, 1-38.

Faddeeva, V. N. Computational Methods for Linear Algebra. New York: Dover Publishers, Inc., 1959.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R., and Gifford, J. A. Developments in latent trait theory: Models, technical issues, and applications. Review of Educational Research, 1978, 48, 467-510.

Kendall, M. G. and Stuart, A. The Advanced Theory of Statistics (Vol. 2). New York: Hafner Publishing Company, 1973.

Lord, F. M. Application to Item Response Theory to Practical Testing Problems. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers, 1980.

Lord, F. M. and Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Likert, R. A technique for the measurement of attitudes. Archives of Psychology, 1932, No.140.

Louis, T. A. Finding the observed information matrix when using the EM algorithm. The Journal of the Royal Statistical Society, Series B, 1982 , 44, 226-233.

Masters, G. N. A Rasch model for rating scales. Doctoral dissertation, University of Chicago, 1980.

Mosteller, F. and Rourke, R. E. Sturdy Statistics. Reading, Mass.: Addison-Wesley Publishing Company, 1973.

Rasch, G. Probablistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Denmarks Paedogogiske Institut, 1960 (Chicago: Uniersity of Chicago Press, 1980).

Samejima, F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement No. 17, 1969.

Samejima, F. A general model for free-response data. Psychometrika Monograph No. 18, 1972.

Samejima, F. Normal ogive model on the continuous response level in the multidimensional latent trait space. Psychometrika, 1974, 30, 111-121.

Stroud, A. H. and Secrest, D. Gaussian Quadrature Formulas. Englewood Cliffs, N.J.: Pretice-Hall, 1966.

Thissen, D. Marginal maximum likelihood estimation for the one-parameter logistic model. Psychometrika, 1982, 47, 175-186.

Thurstone, L. L. Multiple Factor Analysis. Chicago: University of Chicago Press, 1947.

ABSTRACT

This study examines the application of the MML-EM algorithm to the parameter estimation problem of the three-parameter normal ogive and logistic polychotomous item response models. A three-parameter normal ogive model, the Graded Response model, has been developed on the basis of Samejima's two-parameter graded response model. A three-parameter logistic model, the Rating Scale model, has been developed based on Andrich's (Andrich, 1978) Rasch polychotomous response model. In the three-parameter models, an item response parameter is resolved into two parameters, namely, the item location and the category (step) parameters. In the case of the Likert-type questionnaire (Likert, 1932), where only a single scale is employed to evoke different responses for test items, the three-parameter model is expected to be more useful in terms of prediction and analysis because the estimates of the slope and location parameters associated with the points on a single Likert scale can be separately estimated. The advantages of this type of model are demonstrated by analyzing actual data sets. An extension of measurement models to analytic models is also discussed. A specific analytic model, the structural item response model, is a development based on the Graded Response model. In the model, the design matrix is incorporated to represent a structure among subject groups.

TM 810 201